MEDIAL PREFRONTAL CORTEX SIGNALS PREDICTION ERRORS ACROSS DOMAINS OF PAIN AND COGNITIVE CONTROL

Andrew Jahn

Submitted to the faculty of the University Graduate School
in partial fulfillment of the requirements for the degree
Doctor of Philosophy
in the Department of Psychological & Brain Sciences,
Indiana University
May 2015

Accepted by the Graduate Faculty, Indiana University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Doctoral Committee

_____

Joshua W. Brown, Ph.D.

_____

Aina Puce, Ph.D.

_____

Sharlene Newman, Ph.D.

_____

Jerome Busemeyer, Ph.D.

April 22nd, 2015

Acknowledgements

I would like to thank my advisor, Josh Brown, for his mentorship and support throughout my studies at Indiana University. I would also like to thank my former labmates Derek Nee, who assisted me in neuroimaging analysis and interpretation, and Rena Fukunaga, who offered invaluable advice during my early years as a graduate student (and still does).

My love and gratitude, as always, goes out to my parents for guiding me to become the person I am today.

Andrew Jahn

MEDIAL PREFRONTAL CORTEX SIGNALS PREDICTION ERRORS ACROSS DOMAINS OF PAIN AND COGNITIVE CONTROL

Previous research suggests that the medial prefrontal cortex (mPFC), and the Anterior Cingulate Cortex (ACC), in particular, is functionally segregated as a cognitive/affective gradient with caudal mPFC associated with cognitive processing, and rostral mPFC associated with affective processing (Bush et al., 2000). However, recent reviews have shown that the ACC is less functionally distinct than was originally thought (Etkin et al., 2011) and a recent meta-analysis has pointed out significant regions of overlap in the ACC in response to negative affect, pain, and cognitive control (Shackman et al., 2011). In light of these findings, one important hypothesis to test is whether the ACC shows a similar functional homogeneity in response to violations of expectation across different domains of potentially aversive or cognitively demanding stimuli.

This dissertation proposes an experiment to test this distinction between cognitive and affective components of the ACC and is motivated by a recent unifying computational model of ACC function, the Prediction of Response Outcomes (PRO) model (Alexander & Brown, 2011). Different theories of ACC function can then be compared to a current model of ACC function attempting to reconcile all of these experimental findings within a framework of the ACC serving as an action-outcome predictor (Alexander and Brown, 2011). According to this model, prediction signals are generated within the ACC which are then compared and evaluated against the actual outcome that is received, a framework that accounts for a wide variety of error-related and reinforcement learning effects found in the literature.

Here we will explore whether and how the mPFC, and in particular the ACC, is involved in the generation of prediction signals across qualitatively different levels of aversion, as there

has not yet been a study that has combined both violations of expectation of a predicted level of pain as compared to the violation of a predicted level of required cognitive control. The studies reported here provide a background for possible hypotheses for how the ACC may respond to prediction of different levels of aversive outcomes.

_____

Joshua W. Brown, Ph.D.

_____

Aina Puce, Ph.D.

_____

Sharlene Newman, Ph.D.

_____

Jerome Busemeyer, Ph.D.

Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

1.1. Thesis Structure

This dissertation consists of 5 chapters which cover in total 3 functional neuroimaging (fMRI) studies exploring competing computational models of ACC function. Chapter 1 provides a conceptual overview of one of the most researched structures of the medial prefrontal cortex, the anterior cingulate cortex (ACC), along with a comparison of different computational models. This chapter also provides the rationale and competing hypotheses for the present study, along with a review of neurological reward signals in clinical populations and how current computational models relate to these findings. Chapter 2 details an fMRI study to compare two leading theories of ACC function, the conflict monitoring model and the PRO model. Chapter 3 outlines a study where regressors directly generated from the PRO model are applied to fMRI data to distinguish between prediction- and outcome-related regions of the cingulate cortex. Chapter 4 compares predictions from the reinforcement learning model against prediction from the PRO model about how the mPFC processes prediction error across cognitive and affective modalities. Finally, Chapter 5 is a general discussion tying together common threads throughout the trio of neuroimaging studies discussed in this dissertation, and how they shed light on the role of the mPFC as a respone-outcome predictor.

1.2. Overview of the ACC

The ACC is an extensive cortical region closely situated to and sharing several reciprocal connections with nearby motor and premotor areas, as well as areas of the limbic system near the pregenual region (Pandya, Van Hoesen, & Mesulam, 1981; Pickard & Strick, 1996). Due of its

position in the middle of the brain, the ACC receives connections from and extends connections to several different limbic, thalamic, and cortical areas involved in cognitive control and resolution of conflict (e.g., the dorsolateral prefrontal corex, or DLPFC; Holroyd & Coles, 2002). In addition, the ACC receives a wealth of information from sensory areas, and in turn projects dense bundles of efferent fibers to cortical areas involved in motor responses (Pickard & Strick, 1996). Several neurophysiological studies have suggested that the ACC is involved in the control of motor, and especially hand, movements; the detection and resolution of conflict; and processing modulatory dopaminergic signals from midbrain areas which influences the amount of ACC activity under conditions of stress and arousal (Paus, 2001). In addition, neuroimaging studies have revealed several distinct functional roles for the ACC, such as monitoring competition between expected value and risk (Alexander & Brown, 2010), error detection and processing (Gehring et al, 1993; Gemba et al, 1986), response conflict (Botvinick et al, 1999; MacDonald et al, 2000), error likelihood (Brown & Braver, 2005), and prediction of response-outcome associations (Alexander & Brown, 2011).

Subdivisions within the ACC

Several distinct subregions of the ACC have also been delineated, with a traditional cognitive/emotional model segregating the dorsal, "cognitive" region of the ACC (dACC; BA 32 & 24) from the more ventral, "emotional" region inferior to the genu of the corpus callosum (BA 25) (Bush et al., 2000). This dichotomy has persisted over the past decade, owing to the perceived functional similarity between pregenual ACC and limbic areas, and the physical proximity of the dorsal ACC and regions involved in the implementation of cognitive control, such as the DLPFC. Recent research has challenged this idea however, and meta-analyses have

shown that similar regions of dACC are recruited for both cognitive and emotional processing (Etkin et al., 2011; Shackman et al, 2011).

Therefore, a testable hypothesis would be to observe whether the generation of prediction signals – a mental process that would fall within the "cognitive" domain of the traditional cognitive/emotional model – would recruit similar populations of neurons for the prediction of pain as opposed to the prediction for the necessity of a high level of cognitive control. The recruitment of similar populations of neurons for predicting outcomes across both domains would lend support to the hypothesis that the dACC is involved in the generation of prediction signals more generally, regardless of whether it involves a more emotional (as indexed by a measure of arousal, such as galvanic skin response or increased heart rate) as opposed to a more cognitive outcome. Testing this would also need to take into account differences in cognitive processing associated with hemispheric morphology (Amiez and Petrides, 2012; Amiez et al., 2013), and also regional ACC differences in conflict, switching, and error detection (Nee et al., 2011).

The ACC lies immediately ventral to the supplementary eye fields, a cortical area involved in the processing of eye-related movements, and shares reciprocal connections with this area (Schall, Stuphorn, & Brown, 2002). Furthermore, the ACC synapses onto the superior colliculus, a midbrain structure serving as a key hub in the processing and relaying of visual information from the optic nerves to the visual cortex (Leichnetz et al, 1981). In addition to these visual-related areas, the ACC shares a rich set of connections to motor-related areas, such as the pre-SMA and motor areas (Paus et al, 2001) and shows anatomically distinct activation patterns associated with performance to different types of motor tasks. For example, more anterior regions of ACC activate in response to task-related responses involving speech or eye

movements, while the posterior zone of the cingulate is responsive to hand-based responses (Paus et al, 1993). The location of the anterior cingulate is therefore well-situated to process incoming visual stimulation and to evaluate and choose among appropriate motor actions in response to external stimuli, based on past reinforced or punished behavior. This has led to a theory of the ACC as serving as an evaluation mechanism for establishing action-outcome associations, whereas nearby pre-SMA cortical regions receive input from the ACC in order to select appropriate action sets (Rushworth et al, 2004). This framework has spawned several influential theories of how the anterior cingulate detects salient information in the environment, such as errors, and then uses this information to update future actions.

Theories of ACC Function

As the ACC encompasses a large cortical area, its observed functions are likewise diverse: The ACC has been shown to be involved in detecting errors (Falkenstein et al., 1991; Gehring et al., 1993), the implementation of executive function in order to modify behavior in order to more effectively interact with one's environment (Miller and Cohen, 2001), monitoring conflict (Botvinick, Cohen, & Carter, 2004; Carter et al., 1998), processing the likelihood of committing an error (Brown & Braver, 2005), and the experience of pain, both directly by the organism and indirectly through observing others experience pain (Lamm et al., 2011). Error detection and cognitive control, in particular, have been cognitive processes most reliably associated with ACC activation, both for discrepancies between intended and actual responses (Scheffers and Coles, 2000) and discrepancies between intended and actual outcomes (Holroyd and Coles, 2002).

Each of these theories will be examined in turn, as will a review of the key studies lending support to each of these theories. In addition, lesion studies will be discussed which

provide further insight into the necessity and functional specificity of the medial PFC in each of these processes. Finally, potential confounds with each of these studies, such as infrequency and RT, will be analyzed and discussed within the framework of the PRO model (Alexander & Brown, 2011). This background will provide the necessary context for comparing these competing theories against each other to provide the best interpretation of the results of the main experiment discussed in this dissertation.

Error Detection

Early studies of the ACC revealed that a critical functional role of this area was in responding to both the commission of errors and receiving error feedback. Furthermore, given data suggesting that the ACC was involved in attentional processes, investigators hypothesized that this region was involved in the maintenance of focus in order to prevent the commission of errors, and therefore was especially sensitive to sensory information indicating that an error had occurred (Badgaiyan & Posner, 1998).

One model posited that the ACC acted as a comparator through the evaluation of present states to past states, and used that information to detect whether an error occurred or not. Early monkey neurophysiology studies detected negatively deflected error field potentials within the anterior cingulate during a motor learning task when primates initiated incorrect movements (Gemba et al, 1986). Follow-up studies using electroencephalography (EEG) in human subjects revealed that electrodes placed on the anterior frontal scalp located above the medial PFC showed a greater negative deflection, or error related negativity (ERN), both when the subject committed an error and when the subject received feedback indicating that an error had occurred. This latter occurrence is referred to as feedback error related negativity (fERN; Holroyd &

Coles, 2002), in order to distinguish it from the mere commission of an error (Falkenstein et al, 1991). The ERN was shown to occur during the time of actual error commission although no feedback about the response had yet been processed (Gehring et al., 1993). Furthermore, the amplitude of the ERN correlated inversely with the strength of responding on the next trial; i.e., a larger ERN led to a reduction in the strength of a squeeze necessary to make a response. Critically, the size of the ERN also positively correlated with the probability of making a correct response on the next trial, as well as increasing the RT on the next trial.

Individual differences were shown to have a significant effect on the ERN, depending on both contextual and personality factors. For example, the size of the ERN was augmented when subjects were told to emphasize accuracy over speed, which in turn led to investigations of the effects of individual differences on the ERN (Gehring et al, 1993). Specifically, trait measures such as conscientiousness were shown to negatively correlate with the size of the ERN, while personality measures of neuroticism were shown to positively correlate with the size of the ERN (Pailing & Segalowitz, 2004). These findings of individual differences have been extended to other neuroimaging modalities such as fMRI, which have shown that self-report measures such as mind-wandering and absentmindedness have been negatively correlated with activation profiles in the ACC in response to errors, suggesting that less attentive subjects are not engaged in the on-line evaluation of the consequences of actions that lead to errors (Hester, Fassbender, & Garavan, 2004). Taken together, these experiments suggest that error commission leads to a signaling for an increased need for vigilance on succeeding trials and a widening of attention to focus on task-relevant features of the stimulus – such as the color of the word when performing a Stroop task – and that neural activation in response to errors can be modulated by individual differences.

Additional EEG studies utilizing source localization procedures suggested that the ERN is generated within the ACC and/or supplementary motor area, and furthermore is elicited not only in response only to the commission of errors, but instead is modulated by the nature of the error. For example, in a study by Dehaene et al (1994), brain electrical source analysis (BESA) was used to model dipoles of neural activity in response to error commission, and the majority of the variance attributed to this signal was found to correspond to the anterior cingulate. In addition, the authors found significant differences between two types of errors: Slips and mistakes. Slips were defined as incorrect responses that were made consciously and of which the subject knew were incorrect after commission; mistakes, on the other hand, were defined as incorrect responses made on the basis of faulty knowledge, such as misinformation or temporarily forgetting which correct action is mapped onto which specific motor response. Similar EEG studies investigating the neural correlates of error detection have shown that that slips elicited larger ERNs than mistakes (Dehaene et al, 1994), lending support to the hypothesis that the anterior cingulate is responsible for the on-line monitoring of performance when the subject is aware of the contingencies of their response.

A further delineation of heterogeneous ACC involvement in error detection was found by Luu et al (2003), where a sizeable time delay allowed the temporal dissociation between error commission and error feedback. The dorsomedial ACC was found to be responsive to both the ERN and feedback-related negativity, while the rostromedial ACC appeared to be more specific only to the ERN.

Conflict Monitoring

In light of these findings, fMRI experiments began to test whether this comparator hypothesis could account for a more general range of ACC activation in different contexts, including heightened levels of activation in response to conflicting responses that were incompatible. If the comparator model were true, then ACC activation should be greater for conditions in which the subject committed an error, rather than in conditions when high levels of conflict were present. However, researchers found the opposite: Higher ACC activation was present when there was conflict between competing and incompatible responses, regardless of whether a correct or incorrect response was made. This led the investigators to conclude that the ACC appears to be involved more in the monitoring of conflict, as indexed by greater RT, rather than errors per se (Carter et al., 1998). This led to the creation of theoretical framework in which the ACC acts more as a monitor of performance, rather than a comparator that evaluates a current state against a past state (e.g., whether the current state is correct or not, given information about a previous state).

Furthermore, the ACC was found to be preferentially activated to conflict between actions as opposed to conflict between stimuli. For example, in a study by Van Veen et al (2001), a flanker paradigm was used to distinguish between stimulus-incongruent and response-incongruent conditions. In the Stimulus-Incongruent (S-I) condition, the center stimulus was incongruent but mapped onto the same response as the distracting flankers, while in the response-incongruent (R-I) condition, the center stimulus was mapped onto a different response as the distracting flankers. After controlling for reaction time by comparing equivalent RTs in the S-I condition to equivalent RTs in the R-I condition, greater activation in the ACC was

shown for R-I condition as opposed to S-I condition, lending support to the hypothesis that this region is preferentially involved in response incongruency (Van Veen et al, 2001).

As a result of these findings, a model of ACC activation emerged which was predicated on the hypothesis that the ACC acts as an evaluation mechanism which specifically detects the presence of conflict (Botvinick et al, 2001). The model suggested that by detecting the presence of conflict between competing and incompatible responses, the ACC in turn communicates with other prefrontal areas, especially the DLPFC, in order to facilitate the implementation of cognitive control. This model accounted for speed-accuracy tradeoffs as well, with model simulations showing that trials after periods of low conflict evinced lower RT and higher error rates on the next trial, with the opposite pattern holding for high-conflict trials. These simulations suggest that the correlation of ACC activation with greater RT signifies that the presence of conflict elicits ACC activity, which signals the need for greater cognitive control (Yeung & Cohen, 2004).

However, it is important to note that ACC activation is not identical across all situations in which response conflict is present, and instead is dependent on contextual factors. In a study by Botvinick and colleagues (1999), the authors set out to test whether selection-for-action – the focusing of attention on relevant sensory input – was a stronger account of ACC activation than conflict. To that end, RT was measured in response to an incongruent as compared to a congruent stimulus as an index of the conflict present during a particular trial. To test these competing hypotheses, a flanker task was implemented in which different combinations of sequential trials were compared. For example, an incongruent trial following an incongruent trial (iI) would exhibit the greatest selection-for-action related activity, as the previous incongruent trial would prompt more attentional focus on relevant stimuli. By contrast, it was hypothesized

that incongruent trials following congruent trials (cI) would exhibit the greatest amount of conflict-related activity, as subjects are relatively unprepared for a cI trial as compared to iI or iC trials. The authors found that cI, as compared to all other possible trial combinations (cC, iC, and iI) exhibited the greatest amount of ACC activity. Furthermore, a significant correlation was observed between RT and the strength of activation observed in the ACC. These results were consistent with the hypothesis that the ACC acts as a conflict-monitoring mechanism (Botvinick et al, 1999).

These studies led to conflict monitoring becoming the dominant model of ACC function for several years. Subsequent experiments designed to test the effect of conflicting responses during performance have also elicited reliable and robust ACC activation. For example, in a study by Dreher & Grafman (2003), dual-task performance was contrasted against a condition involving task-switching, with participants engaged in the dual-task condition showing greater rostal ACC (rACC) activity as compared to task-switching. Their interpretation of the rACC activation in response to this contrast was that the same populations of neurons in the rACC are recruited to evaluate both potential motor responses, which can come into conflict. However, whether the rACC activation was due to two motor responses, or the conflict of the responses, or both, remains to be resolved.

Primate vs. Human Results

However, new results arose to challenge this dominant viewpoint – in particular, inconsistent primate neurophysiological data, and studies focusing on reaction time and the probability of outcomes. In addition to the plethora of theories surrounding ACC function, neurophysiological work involving primates has contributed finer-grained data about cingulate function at the single-

unit level. However, these results must be approached with caution as well. Although the primate brain is similar to the human brain, important differences exist. For example, a dorsal-caudal extension of the cingulate, area 32', exists only within the human brain (Rushworth et al, 2004). In addition, there is considerable variability within the cingulate of humans themselves, with 30-50% of individuals having a cingulate divided in half by a sulcus in at least one hemisphere (Cole et al, 2009). This paracingulate sulcus has also been shown to have a pronounced association with different activity profiles in response to feedback-related error activity (Amiez et al, 2013). Furthermore, the methodology of neurophysiology is more localized than EEG and fMRI methods, leading to divergences in localizing effects in certain regions of the brain (Cole et al, 2009). Although field potential studies have been carried out in the human ACC (e.g., Lenz et al, 1998) and have been useful in providing both high spatial and temporal resolution, often they are conducted only in patients presenting with abnormalities such as seizures (Wang et al, 2005).

With these anatomical and methodological caveats in mind, however, one of the most salient differences between monkey neurophysiological data and neuroimaging studies is that conflict effects, although heavily documented in EEG and fMRI literature (Botvinick et al, 2001), have been relatively inconsistent in being found within the ACC proper, with conflict effects found instead in the pre-SMA (Cole et al, 2009). Indeed, one primate study teasing apart contributions of error, reward, and conflict in a saccade countermanding task failed to find any individual neurons responsive to conflict, as defined by the interruption of a prepared saccade (Ito et al, 2003). However, effects of error and reinforcement were observed in roughly equal proportions – including, interestingly, neurons responsive to the omission of an expected reward. Even studies which have putatively reported conflict-related effects within the cingulate may be confounded by differences in error frequency between conditions (Cole et al, 2009), and other

primate studies have found conflict-related effects in primates with ACC lesions (Mansouri et al, 2007).

Although it is possible that conflict effects may be located more dorsal to the cingulate and closer to the pre-SMA and supplementary eye fields (Nakamura et al, 2005; Isoda & Hikosaka, 2007), a more parsimonious explanation is that the cingulate performs a process other than conflict monitoring. For example, recent studies have found that the dorsal ACC processes unsigned prediction error – in other words, the unexpectedness of a result, regardless of valence (Hayden et al, 2011) – and other cell recordings within the mPFC have shown greater neural spiking rates to error trials when a reward was expected, possibly representing the unexpected non-occurrence of an event (Amador et al, 2003; Alexander & Brown, 2011). The ramifications of these primate studies will be explored more in-depth when discussing the error likelihood and PRO models of ACC function (Brown & Braver, 2005; Alexander & Brown, 2011).

ACC and Reaction Time

An additional challenge to the conflict monitoring model arose from the fact that higher RT, traditionally used as an index for the amount of cognitive control or conflict present on a given trial, is highly correlated with ACC activity. In a series of simulations run by Grinband and colleagues (2011), the authors found that by binning the RT of trials into quantiles and covarying out the effects of RT, ACC-related effects disappeared. A similar study carried out by Carp and colleagues (2010) matched congruent and incongruent trials on RT and contrasted the mean neural activity present in each of these conditions. mPFC and ACC effects disappeared after controlling for RT in this way, although matching error trials with correct trials with equivalent RTs still elicited heightened levels of ACC activation. The authors suggested that the conflict

monitoring effects shown in other studies may be driven more by error commission rather than by effects of congruency and incompatible responses. Finally, a meta-analysis conducted by Yarkoni et al (2009) showed that, across a wide range of tasks purportedly involving cognitive control, increases in RT correlated linearly with activation within the pre-SMA and anterior cingulate areas, irrespective of what task the participants were engaged in. Taken together, these recent studies have lent significant support to the hypothesis that the ACC, acting as an attentional mechanism, may be particularly sensitive to time on task, which explains a larger amount of the variance of activity in this area as compared to the nature of the task they are engaged in.

However, while RT has indeed been shown to be significantly correlated with ACC activation, it is by no means the only factor driving neural activity in this area. For example, a study by Nee & Jonides (2008) comparing negative priming (i.e., instructing the participant to ignore a specific stimulus) to proactive interference (i.e., instructing the participant to forget a specific stimulus) found no significant ACC activation as a result of this contrast, despite the fact that there were significant RT differences between the two conditions. Furthermore, a study by Nelson and colleagues (2003) directly compared conditions involving response conflict to conditions involving familiarity conflict, in which a cue had been present on previous trials, and therefore entailed a high feeling of familiarity, but was not present on the current trial and therefore required a rejection response. The investigators found a double dissociation between response conflict and familiarity conflict, with the former eliciting a canonical conflict response in the medial PFC and familiarity conflict eliciting activation within the left inferior frontal gyrus, even though RTs for both conditions were nearly identical. Both of these studies suggest

that mPFC activation may be sensitive to psychological context rather than simply correlation with higher RTs.

ACC and Pain

In addition to situations involving high degrees of conflict, the anterior cingulate is also highly responsive to aversive stimuli, and especially to the perception of pain. Interestingly, it is not merely the presence or intensity of pain that elicits neural firing in this area, but rather the perceived unpleasantness of pain. In a study by Rainville et al (1997), participants were subject to hypnotic suggestion that either attempted to increase or decrease the perceived unpleasantness of submerging their hand in hot water. The authors found that anterior cingulate activity increased dramatically in response to hypnotic suggestion which amplified their subjective perception of the unpleasant stimulus, as compared to a hypnotic condition designed to decrease the subjective unpleasantness of the stimuli. However, across both hypnotic conditions, activation in the somatosensory cortex was nearly identical. This study suggests that the ACC, which shares many connections with the somatosensory cortices, is not exclusively responsive to aversive stimuli, but rather to the individual's subjective response to aversive stimuli.

Based on the anterior cingulate's supposed role in processing aversive stimuli, an experiment by Derbyshire et al (1998) attempted to dissociate the ACC's response to painful stimuli and Stroop stimuli. It is plausible that the ACC is more involved in attentional processes rather than pain per se, and therefore that the attention-related processes of pain and Stroop ktask would lead to a high degree of overlap of activation within the ACC. However, the authors found distinct subregions of the ACC responsive to each condition, with painful stimuli recruiting more anterior portions of the ACC and incongruent Stroop stimuli recruiting more caudal regions of

the ACC. This dissociation between the processing of pain and processing of tasks involving cognitive control may suggest that a similar functional dissociation is present when generating prediction signals for each of those conditions. However, this study did not include auxiliary measures of arousal, such as pupil dilation or GSR, which could be important confounds in interpreting cortical responses to aversive stimuli.

Lesion Studies

Although the above neuroimaging literature has implicated the dACC as playing a critical role in the signaling for cognitive control when necessary, the most direct test of a brain structure's necessity in a cognitive process is through lesion studies. For example, if it can be demonstrated that a subject without an ACC still performs equivalent to controls on tasks involving cognitive control, then that would argue against the necessity of that area's involvement in the hypothesized cognitive process. Studies involving human subjects with lesions are relatively rare and suffer from low power, but can still reveal important aspects of neural functioning.

The ACC, in particular, has been the subject of several lesion studies that have shown conflicting and counterintuitive results. For example, a single-subject lesion study of a patient with left ACC damage exhibited both smaller ERNs and increased RT in response to incongruent stimuli in a spatial Stroop paradigm. This study showed that conflict monitoring and error detection, at least in this patient, do not both come from the same area of ACC, suggesting that these processes occur in different areas. However, while the ERN was shown to be attenuated in the patient, the conflict response (a waveform called the N450) was actually enhanced (Swick & Turken, 2002). This suggests that conflict monitoring occurs in a nearby prefrontal area, such as the DLPFC, before information about the conflict is sent to the ACC.

On the other hand, a lesion study conducted by Fellows & Farah (2005) compared the performance of individuals with dACC lesions to that of controls across a battery of tasks hypothesized to involve cognitive control. These tasks included a Stroop task and a go-nogo task which are known to elicit significantly greater increases in RT after errors, and to induce significantly greater amounts of errors during incongruent trials. The results showed no significant interactions between group and task, suggesting that the dACC is not necessary for the implementation of cognitive control. Furthermore, the authors pointed out that tasks involving cognitive control may be confounded with emotional responding, which in turn could simply be associated with the ACC's involvement in regulating muscle tone. In any case, it is apparent that although this structure is somehow associated with cognitive control, it is not strictly necessary for it.

In sum, these lesion studies suggest that the dACC may not be indispensable for signaling the DLPFC to implement cognitive control. An alternative explanation may be that patients with ACC lesions are usually ipsilateral, and that furthermore they may be compensating for required cognitive control by recruiting nearby cortical areas. However, two lines of evidence argue against this interpretation. First, one of the lesion patients examined in the Fellows & Farah (2005) had extensive medial ACC damage encompassing dACC bilaterally, but showed a similar pattern of error rates and RT difference between congruent and incongruent conditions as did the other lesion patients and the control group. Secondly, lesion studies of other areas of the brain – such as the orbitofrontal cortex – have shown that those regions appear to be specific to the cognitive processes they are hypothesized to be involved in. For example, patients with OFC lesions exhibit significantly impaired performance in decision-making tasks such as the Iowa Gambling Task and Wisconsin Card Sorting Task, as well as decreased autonomic activity in

response to highly risky gambles (Bechara et al, 1994). Even though the patients in this study had suffered from their lesions for a comparable amount of time as the lesion subjects in the Fellows & Farah (2005) study, there was no evidence of recruitment of other cortical areas in order to support their deficits in decision-making.

Furthermore, although these lesion studies have shown no significant differences in error rates between the lesion patients and controls, other experiments have revealed that patients with ACC damage are less likely to correct for their mistakes on trials immediately following an error. For example, patients with ACC lesions are less likely to be aware that an error has occurred (Swick & Turken, 2002), and in a double-dissociation lesion study, patients with ACC damage were found to have impaired rates of error correction but not error suppression, while patients with basal ganglia lesions showed the opposite pattern of impairment (Hochman et al, 2015). These results suggest that there may be a necessary role for the ACC for the actual detection of errors, which would be consistent with the hypothesis that this area is involved in the comparison of actions against their predicted outcomes. How lesions affect the transfer of information from the ACC to the DLPFC and other cortical regions supposedly involved in the implementation of cognitive control, however, is less well understood.

The ACC, Negative Reinforcement Learning, and Error Likelihood

Another influential theory of ACC function, the negative reinforcement learning model (Holroyd & Coles, 2002), has attempted to unify several of these different findings by casting the ACC as processing the valence of outcomes. According to this theoretical framework, the cingulate assigns more weight to negatively valenced events than positively valenced events, thus driving learning effects – explaining, for example, why certain regions of the cingulate show greater

activity in response to monetary punishment as opposed to monetary reward (Knutson et al, 2000). The proposed mechanism for this learning is mesencephalic dopamine projections to the mPFC, which show higher phasic activity when events are better than expected, and phasic decreases when events are worse than expected (Schultz et al, 1997). The authors hypothesized that dopamine should exert a dampening effect on ACC activity, and that when this dopamine release is decreased or otherwise inhibited, ACC activity levels show a corresponding increase.

However, one confound of this model, in addition to the error detection (Gehring et al, 1993) and conflict monitoring (Botvinick et al, 2001) models, is that error-related events are usually infrequent, and therefore observed ACC activation in response to errors may be due to infrequency effects. For example, in a study by Jessup et al (2010), the authors tested this hypothesis by presenting subjects with an experimental paradigm in which errors were relatively more frequent than rewarding outcomes, similar to the situation of playing the lottery. The authors found that participants exhibited greater ACC activation in response to correct outcomes in the same region typically associated with error-related activity, suggesting that this area is more sensitive to infrequency effects than errors themselves.

Studies using other modalities have found similar results. For example, an EEG study by Ferdinand et al (2012) matched positively and negatively valenced events on frequency, and found statistically equal amounts of feedback related negativity (fRN) from electrode sites placed over the anterior frontal scalp. A follow-up study by Garofalo et al (2014) also used EEG, but examined the fRN in response to the presence or absence of an electrical shock. By pairing a stimulus with a high percentage of receiving an electrical shock, the authors were able to examine trials where the predicted shock failed to occur. Similar to the Ferdinand et al 2012

paper, higher amounts of feedback related negativity were found over anteriorfrontal electrode sites located over the mPFC.

These results were consistent with earlier studies examining the response of the ACC to error likelihood, in which participants were presented with cues signifying the probability of a switch in the change signal delay paradigm (Brown & Braver, 2005). In this paradigm, participants are presented with an arrow pointing in one direction, and have a response mapped onto the direction of that arrow. However, within a subset of trials, an arrow will appear on the screen pointing in the opposite direction, signifying that the prepotent response must be overridden and the opposite response chosen. During these "Change" trials, therefore, there is a higher likelihood of committing an error. It was tested whether the presentation of a cue signifying a high probability of receiving a Change trial, as opposed to cues signifying a low probability of receiving a Change trial, would elicit different patterns of activity in regions of medial PFC known to be involved in evaluative processes. A contrast of these high error likelihood cues as compared to low error likelihood cues revealed activation in the dorsal ACC, suggesting that this area is involved in processing the likelihood of committing an error, rather than conflict itself. Given these results, a plausible interpretation of the conflict processing literature is that higher levels of conflict entail a greater likelihood of error commission, a scenario that is encompassed within the error likelihood model (Brown & Braver, 2005).

Further refinements of the error likelihood model led to a theoretical framework in which the ACC was hypothesized to be more generally involved in the prediction and evaluation of the outcomes of one's actions (Alexander and Brown, 2011; Brown and Braver, 2005; Magno et al., 2006). In support of this hypothesis, while conflict monitoring studies have implicated the ACC as a key hub in a prefrontal network involved in detecting high-conflict states and the subsequent

recruitment of the DLPFC for the implementation of cognitive control (Kerns et al., 2004; MacDonald, 2000), ACC activation was observed in response to the mere imagination of error outcomes, apart from any overt motor response (Jahn et al., 2011). Furthermore, recent studies have shown that the ACC is not only responsive to the actual commission of errors and receiving error feedback, but also in generating prediction signals about future scenarios in which errors are likely (Aarts et al, 2008). For example, regions of the ACC and pre-SMA were found to activate in response to cues predicting the probability of receiving an incongruent Stroop stimulus, an outcome which would require the recruitment of cognitive control (Aarts & Roelofs, 2011).

The ACC as an Action-Outcome Predictor

These results point toward the ACC's role as an action-outcome predictor, which generates simulations about possible future states associated with the execution of specific actions. This is a core concept of the PRO model, where the ACC, and the mPFC in general, is posited to be involved in learning response-outcome associations in specific stimulus contexts. In support of this, a recent neuroimaging study by Jahn et al (2011) found that the ACC was activated in response to errors that were merely imagined, and that this area of activation recruited a similar area of cortex responsive to the feedback of actual errors. Furthermore, a model-based fMRI study found that regressors generated by the PRO model tessellated the ACC into distinct prediction and evaluation regions, with the more medial ACC associated with evaluating outcomes, and more posterior and anterior regions of the ACC associated with prediction processes (Jahn et al, 2014). This latter study, in particular, was the first to directly use model-

based regressors from the PRO model to examine their loading on ACC activity. The ramifications of this model and its relation to clinical populations is discussed in the next section.

1.3 Prediction-Related Signals of Avoidance Motivation in Clinical Populations

Current literature suggests that individuals suffering from drug addiction are more likely to be motivated by reward signals when beginning the consumption of a drug (i.e., seeking a drug high), and more likely to be motivated by the avoidance of the negative consequences of withdrawal after developing a high tolerance to the drug (i.e., the negative reinforcement theory of addiction; Ahmed & Koob, 2005; Koob & Lemoal, 2005). In order to distinguish between the neural mechanisms of these two (non-exclusive) possible bases of drug addiction, existing decision-making literature regarding drug addiction in both behavioral and neuroimaging settings will be reviewed. Here, in relation to cognitive models of learning and approach behavior, particular attention will be given to how individuals with drug addiction respond to predictive cues for drug administration as opposed to cues predictive of qualitatively different rewarding stimuli, such as monetary reward.

Next, a neurobiologically plausible network will be discussed that is involved in the generation of these prediction signals as well as the evaluation of outcome, and how these circuits differ in drug populations as compared to controls. In particular, the roles of several cortical regions involved in reward processing and prediction signals, including the anterior insula, anterior cingulate cortex, and subcortical structures including the dorsal striatum and nucleus accumbens, will be examined, as well the interactions between these regions in processing evaluation and reward signals. A deeper understanding will provide insight into how

this network operates in healthy controls, as well as how it can potentially fail to function appropriately in substance-dependent populations.

Approach vs. Avoidance Theories of Motivation

Classical theories of drug addiction posit that the decision to take drugs is motivated by the distinct drives of approach and avoidance, depending on whether the individual is motivated to seek a drug high or avoid the effects of drug withdrawal. Approach behavior has been theorized to be driven by positive associations with drugs and drug cues, such as drug paraphernalia and environments in which the drug is taken (Stewart et al, 1984). Avoidance behavior, on the other hand, is hypothesized to be driven by an avoidance of the negative effects of drug withdrawal (Siegel, 1999). For example, a cocaine user who has not taken the drug for an extended period of time and is beginning to suffer from an initial period of withdrawal may be more motivated to take the drug in order to alleviate aversive symptoms.

Gray's BIS/BAS Model

The duality of approach and avoidance behavior can be summarized within a single theoretical framework put forth by Gray (1970), called the Behavioral Inhibition System / Behavioral Activation System (BIS/BAS; Kumari et al., 1996). This theoretical construct contains two polar motivational drives that are responsible for avoidance and approach behavior, and has become a particularly influential model of risk behavior, and by extension the use of and experimentation with drugs. Studies testing the internal validity and convergent validity of the measures have shown that higher self-report ratings on the BIS in general are associated with higher levels of neuroticism and nervousness, while higher ratings on the BAS are associated with greater self-

report measures of happiness (Carver & White, 1994). In particular, the BAS, associated with higher levels of approach behavior, has been correlated with drug use (Franken et al, 2006) and cravings to avoid the negative symptoms of alcohol withdrawal (Franken, 2002)

However, the BIS/BAS scale has also come under scrutiny for showing significantly high correlations with a broad range of psychopathologies, including depression, bipolar disorder, drug addiction, and anxiety (Johnson et al, 2003). Thus, the explanatory power of this scale is relatively low when attempting to distinguish between different subtypes of psychological disorders, including drug addiction. A stronger theoretical framework of approach and avoidance behaviors in drug addict populations, therefore, would be both useful and necessary when attempting to categorize drug-seeking behavior and to discriminate what aspects of approach and avoidance behavior are associated with what aspects of drug addiction. The following section highlights key cortical regions involved in drug addiction and reward processing more generally, in the attempt to create a neurobiologically plausible model of drug addiction focusing on nicotine.

Neurobiological Models of Drug Addiction

A deeper understanding of the neural mechanisms underlying drug addiction involves reviewing the literature concerning how individuals make decisions, what neural mechanisms are involved in these decision-making processes, and how these processes break down or are dysfunctional in individuals with drug addiction. Two prominent theories of decision-making are model-free and model-based reinforcement learning, which posit different cognitive and neurological mechanisms involved in deliberative and habitual behavior, respectively. These models will be compared below and then extended to theories of drug addiction.

Brain Regions Involved in Reward and Decision-Making

In order to formulate a plausible neurobiological model of drug addiction and potential dysfunction in this network, a review of the key brain regions involved in reward and decision-making behavior is necessary. Many of the structures discussed below serve multifaceted functions that only incidentally include decision-making behavior, while other regions appear to be much more specific to reward processing and the decision-making process. However, each is linked by significant differences between drug addict populations and healthy controls in both evaluating the risk involved in taking drugs, and the evaluation of the reward itself derived from drug administration. Each of these regions is discussed in turn, as well as any structural or functional connectivity between them.

Anterior Insula

The anterior insula (AI), a cortical region encapsulated within the lateral areas of the brain between the frontal and temporal lobes, has been shown to play a key role in the monitoring of one's own interoceptive state, including the processing of sensations of pain (Baliki et al, 2009), disgust (Jabbi et al, 2008), and negative stimuli (Critchley et al, 2004). Recent studies have also shown the AI to be a critical mediator of drug-related behavior. In a study conducted by Naqiv et al (2007), smokers who presented with lesions to either the left or right insula were compared to smokers without brain damage. The authors found that the smokers with lesions were more likely to be able to quit smoking and less likely to relapse, as compared to the control smokers. Furthermore, the right insula, as opposed to the left insula, appeared to account for more of the variance in quitting smoking, although lesions to either the left or right insula were more highly

predictive of quitting smoking than lesions to any other area in the brain. This suggests that the insula, through its role in moderating conscious urges and cravings, is a critical cortical substrate for the maintenance of addictive behaviors.

However, it should be pointed out that the presented lesions typically encompassed nearby cortical and subcortical regions as well – including areas such as the putamen and dorsal striatum – which are also involved in habit learning. Although a follow-up analysis revealed that the disruption in smoking behavior was particular to insula regions, the relatively small sample size present in this study, as well as inferential complications arising from any lesion study – such as the degree of plasticity following a neural insult (Müller & Knight, 2006) – make it difficult to pin down the exact contribution of the insula. In light of these caveats, however, it is interesting to note that the smokers with insula lesions did not report any reduction in cravings for other drives, such as eating. A possible hypothesis to reconcile these findings is that the insula generates craving signals for pleasurable associations that are learned over time, whereas more basic drives, such as eating or drinking, may be served by redundant cortical connections, due to their importance in keeping the organism alive.

Given these findings, one theory of anterior insula function casts the anterior insula as a "limbic sensory cortex," receiving projections from the parabrachial nucleus – a primary integration site for interoceptive information from the rest of the body – and providing the foundation for higher-level emotional awareness (Craig, 2003). This is in contrast to the posterior insula portion of the secondary somatosensory cortex, which processes more basic and visceral sensations such as pain, itch, hunger, and thirst (Craig, 2002, 2003).

It is not surprising, therefore, that the insula reacts strongly to situations involving risk and other autonomically arousing scenarios. In a study of gambles with varying levels of risk

where subjects were able to either accept risky gambles or pass on those gambles, greater levels of insula activity were observed when making a risky decision, as well as being predictive of making a decision associated with greater levels of risk (Xue et al, 2010). On the other hand, however, higher levels of AI activity were observed to be correlated with better decision quality in a related risky decision task, the Iowa Gambling Task (Krawitz et al, 2010). Comparing substance dependent populations to controls, those with substance dependency showed overall lower levels of AI activity in response to messages framing decisions as having potentially positive or negative consequences, and correspondingly lower decision quality when deciding whether to take gambles or not. To reconcile these two findings, it is plausible that the AI signals internal states of arousal, and affects decision-making to the extent that a specific action is decided on. For example, those who are committed to already making a risky decision will show correspondingly higher levels of AI activity, while those sensitive to messages framing a decision as potentially risky will show AI activity profiles reflective of a higher vigilance for making more appropriate decisions.

A related cortical region sharing dense reciprocal connections to the anterior insula is the inferior frontal operculum (IFO). Both the AI and the IFO are usually considered a single cortical unit involved in semantic processing (Friederici et al, 2003) as well as processing interoceptive states, and together comprise the gustatory cortex (Jabbi et al, 2007; Krawitz et al, 2011). As an example of the IFO/AI's role in processing gustatory stimuli, both patches of the anterior insula and IFO were activated in response to observing disgusting facial expressions, ingesting bitter liquids which induced subjective feelings of disgust, and reading vignettes intended to elicit feelings of disgust (Jabbi et al, 2008). Furthermore, lesions to the IFO interfere with the recognition and experience of disgust, suggesting that this region is essential for processing

interoceptive features of disgust (Adolphs et al, 2003; Calder et al, 2000). Taken together, the AI/IFO axis appears to play a critical role in regulating the individual's reaction to stimuli eliciting visceral sensations ranging from disgust to cravings. The interoceptive properties of this area of cortex make it an essential region for the interpretation of visceral feelings in the body, which includes the perception of pain and the perception of cravings, whose dysfunction can lead to addictive behavior.

Anterior Cingulate Cortex

The anterior cingulate cortex (ACC), a patch of cortex lining the medial wall of the brain just above the corpus callosum, is a critical region involved in decision-making behavior, as well as the generation of prediction signals involved in the expectation of response-outcome associations (Alexander & Brown, 2011). Experiments have confirmed the role of this region in conflict monitoring (Botvinick et al, 2001) and error detection (Gehring et al, 1993), as well as predicting the likelihood of receiving an error (Aarts & Roelofs, 2011; Brown & Braver, 2005). As drug addiction involves the repetition of behaviors that the individual may consciously know to be harmful to their health, this could point toward a deficiency in the ACC that leads to an inability to appropriately evaluate risky decisions and their consequences.

Several theories of ACC function touch on its dysfunction in neuropsychiatric and substance-related disorders. For example, studies examining the ERN in substance-dependent populations found an attenuated ERN in response to errors (Franken et al, 2007), as well as lower ERN profiles in persons scoring high in impulsivity, a significant predictor of future substance abuse (Olvet & Hajcak, 2008). Due to the ERN's hypothesized role in error detection (Falkenstein et al, 1999, Schefers et al, 1996), conflict monitoring (Yeung, Cohen, & Botvinick,

2004), and subjective responses to errors (e.g., Bush et al, 2000, Gehring & Willoughby, 2002), several constructs could possibly explain how ACC dysfunction leads to risky decision-making behavior and substance dependence. However, none of these theories have made an explicit link between the hypothesized role of the ACC and resulting substance dependence and abuse.

The PRO model of ACC function, on the other hand, recently addressed this issue directly by examining substance-dependent individuals in an fMRI study. It was found that the profile of ACC activity could be best fit by a concave value function, in which smaller rewards were weighted more heavily than larger ones, qualifying these individuals as "risk-averse" in decision-making nomenclature (Alexander et al, in press). Although this characterization of substance-dependent individuals may seem counterintuitive, it could be explained by small and immediate rewards being more salient and appetitive, while larger value payoffs not seen as worth the effort or time (Alexander et al, in press). This represents a significant step in combining neuroimaging results with modeling to explain drug-seeking behavior in substance-dependent populations.

Another example of the involvement of the ACC in addict populations was conducted by Fishbein et al (2005), in which abstinent drug abusers were compared to healthy controls on a risky decision-making task while undergoing PET scanning. The investigators found that, compared to healthy controls, recovering drug abusers exhibited lower levels of perigenual ACC activity in response to risky decisions that had the potential to yield greater rewards, but also entailed greater penalties. This deficit in risk evaluation could be due not only to the ACC's role in prediction signaling, but also because of its dense interconnections with areas involved in appraising the value of potential rewards, such as the medial orbitofrontal cortex (OFC), as well as connections to subcortical areas, such as the amygdala, which are involved in monitoring the

valence of the stimuli that one is presented with (Cunningham et al, 2010). Faulty connections with any of these regions, as well as any deficits within the regions themselves, could lead to the risky behavior observed in these studies.

Furthermore, the ACC, due to its receiving dopaminergic projections from the midbrain area (Williams & Goldman-Rakic, 1998; Holroyd & Coles, 2002) is situated to process reward-related stimuli and the potential reward resulting from actions (Allman et al, 2001). Indeed, the ACC as a whole receives one of the densest dopaminergic innervations of the entire brain (Paus, 2001). This dovetails with findings related to the competition hypothesis of Alexander & Brown, 2010, where the ACC was found to process competing signals from the error likelihood and expected reward from a decision, doing so in a relatively isolated fashion independent of signals from other brain regions (Alexander & Brown, 2010).

Previous research has shown that both drug intake and high levels of stress increase the sensitivity of dopaminergic neurons, which in turn can override the more adaptive, rational aspects of the decision-making process (Saal et al, 2003). Chronic drug abuse, therefore, may lead to a hypersensitization of the dopamine projections between regions of the rostral ACC – including the perigenual region – and areas of the limbic system rich in dopaminergic receptors and playing a critical role in decision-making, such as the dorsal striatum.

Nucleus Accumbens and Striatum

The striatal region of the brain typically refers to the caudate nucleus and nearby putamen, which wrap around the dorsal aspect of the thalamus, following the underbelly of the corpus callosum and forming a major part of the limbic system. The striatum plays a key role in instrumental conditioning, receiving dense dopaminergic projections from the basal ganglia, as well as

projecting to diverse areas of the cerebral cortex, particularly the rostral anterior cingulate (Paus, 2001). One of the predominant models of striatal function is the actor/critic model put forth by O'Doherty and colleagues (2004). According to this framework, the ventral striatum functions as a critic, evaluating expected outcomes against what actually occurred, and whether this outcome was better or worse than expected. This information is then relayed to the dorsal striatum, which serves as an "actor", implementing a new response-outcome policy based on updated representations supplied by the ventral striatum of expected value for specific actions.

However, possibly the most studied region of the striatum implicated in addiction behavior is the nucleus accumbens (NAc). Nestled ventral to the pregenual tip of the corpus callosum and neighboring the rostral anterior cingulate cortex, the NAc forms the main part of the ventral striatum and receives dopaminergic projections from the ventral tegmental area, a region of the midbrain responsible for sending dopamine efferents to several distinct regions of the brain, but particularly to the striatum and forebrain. The NAc can be divided into two major components: the dorsal core and the ventromedial shell (Chiara et al, 2004). It is the shell, in particular, that has been implicated in increased dopaminergic firing in response to rewarding stimuli, such as palatable food and addictive drugs, and has been shown to moderate lever-pressing behavior in rats when seeking the administration of more drug into the accumbens shell.

In addition, NAc sensitivity and activation has been associated with higher scores on the BAS scale and with less efficient inhibitory dopaminergic activity within the striatum and NAc, implying that persons with a greater predisposition to approach rewarding stimuli may have a correspondingly lower threshold for neural firing in reward-related areas such as the NAc (Dawe et al, 2004). Therefore, the NAc could be a candidate region for greater parametric modulation in response to more appetitive, rewarding aspects of a gamble, such as the potential amount that can

be won from making a riskier decision. Of particular interest would be how drug addiction populations differ from controls in reward-related subcortical regions such as the NAc in evaluating both potential monetary gains and potentially high levels of nicotine administration. These considerations will be outlined in greater detail in Appendix C.

Amygdala

Another component of the limbic system which has received widespread attention in both the emotion and decision-making literature is the amgydala, an almond-shaped bilateral structure which lies underneath the temporal lobes at the tail of the caudate. Traditionally, the amygdala has been implicated in emotion processing, especially that of fear, but this view has been shown to be far too limited; more recent studies have highlighted the role of this region in processing valence more generally, regardless of positivity or negativity, with particular sensitivity to the emotional intensity of the stimulus. Furthermore, individual difference measures such as neuroticism and approach-avoidance personality traits have been shown to be correlated with activity in this area (Cunningham et al, 2010).

A complex structure composed of multiple nuclei (Amunts et al, 2005), different components of the amygdala have been shown to be associated with different aspects of reward processing. Furthermore, the nuclei of the amygdala have complex connections and relationships with prefrontal structures, including the OFC. For example, in a study by Schoenbaum et al (2003) the basolateral complex of the amygdala (ABL) was shown to be involved in encoding associations between neutral cues and outcomes, which in turn were then used by OFC to guide behavior according to specific contexts. The experimenters subjected rats to a learning paradigm in which outcome-expectation neurons within the OFC were recorded while the rats learned

contingencies associated with particular odor cues. Before learning, there were similar levels of firing in OFC neurons in both ABL-lesioned rats and controls. However, after learning, ABL-lesioned rats exhibit far less firing in OFC neurons than controls. Thus, the ABL appears to be crucial for the formation of associations between neutral cues and outcomes.

Within humans, the contribution of the amygdala nuclei to decision-making behavior has been examined ever since lesions of the amygdala in monkeys were shown to lead to unusual and risky behavior, such as approaching objects or animals which would frighten monkeys who had their amygdala still intact (Kluver & Bucy, 1939). However, lesion studies have shown that damage to the amygdala disrupts the generation of galvanic skin conductance after commission of risky decisions and impairs performance on decision-making tasks (Bechara et al, 1994), as well as disrupting cocaine self-administration in rats (Koob, 1999). In addition, the complexity of the connections of the amgydala to the prefrontal areas of the brain (via the ABL) and to other subcortical structures (primarily via the central nucleus) has also precluded a clear explanation of its role in drug-related behavior. Therefore, although experimental designs focusing on whole brain analysis may be able to make some inferences about the functional connectivity and mediation of the amygdala on other cortical and subcortical structures, the ability to determine the relative roles of the subnuclei within the amygdala are much more difficult to measure noninvasively in humans. Future improvements in fMRI resolution may resolve this problem.

Orbitofrontal Cortex

Lastly, the orbitofrontal cortex (OFC) rounds out the midbrain-striatum-prefrontal axis of dopamine signaling and reward processing. Considerable research in human populations has established the OFC as a critical region in evaluating rewards, with the medial OFC exhibiting

selective processing for positive rewards, and the lateral OFC showing heightened sensitivity to punishments and potential losses (Rolls, 2004). Thus, this region's role in decision-making has been widely studied, as have deficits in OFC function which correlate with the commission of maladaptive decisions. For example, in the Iowa Gambling Task paradigm (Bechara et al, 1997), healthy controls learn that specific decks are riskier and lead to higher losses over time and avoid them over the course of the experiment. However, both patients with OFC damage and drug addiction populations such as cocaine users show substantially different performance on the task, with drug addiction populations selecting much more often from the riskier decks even though this behavior entails higher overall losses (Verdejo-Garcia et al., 2007). Neuroimaging studies have shown that deficits in decision-making behavior during this task is associated with higher levels of right OFC activity, suggesting that drug addiction populations may be overly sensitive to the reward aspects of that deck, which may also be complemented by an insensitivity to the higher risk and larger losses associated with choosing the riskier decks (Bolla et al., 2003).

Furthermore, while research on primates has shown both the OFC and nearby basal ganglia to contribute to the expectation and receipt of reward, these investigations have also revealed several key differences between the OFC and the basal ganglia, with the most salient difference occurring between the appraisal of the rewarding stimuli and the processing of the value of motor actions. The OFC appears to be more involved in appraising the potential rewards associated with the stimuli (Rolls, 2004), and nearby regions of ventromedial prefrontal cortex (vmPFC) have been shown to be associated with evaluating experienced vs. stated preferences (e.g., McClure et al, 2004; Plassman et al, 2008). The basal ganglia, on the other hand, are associated with evaluating the value of motor actions that result in either rewards or punishments (O'Doherty et al., 2004). For example, striatal activity appears to be contingent upon actual

receipt of reward, and in response to the preparation and execution of movements that will result in reward. In addition, the OFC appears to be involved in higher-level abstract processing of reward information, with OFC neurons involved in the discrimination between different types of reward (e.g., whether the reward is appetitive for different modalities such as sight or smell), as well as comparing a reward to available alternatives (Tremblay & Schultz, 1999).

Relationship of Prediction-Related Systems to Drug Addiction

One plausible hypothesis about drug addiction already alluded to is that drug addiction populations may form maladaptive or dysfunctional predictions about the consequences of their actions. For example, it is possible that an addict may be aware about the potential adverse effects of drug intake, but may not experience or adequately process visceral states that provide subjective signals that a certain course of action may be maladaptive. In a classic study by Bechara et al (1994) involving patients presenting with orbitofrontal cortex (OFC) damage, subjects were more likely to choose from decks that yielded higher immediate gains, but overall higher losses in the long term. As compared to controls, OFC lesion patients exhibited suppressed galvanic skin response (GSR) activity in response to selecting from the risky deck. Since GSR is a reliable measure of physiological arousal, the authors hypothesized that OFC lesion patients were unable to interpret bodily responses appropriately to dissuade them from choosing from overly risky decks.

In addition to the OFC, the anterior cingulate cortex (ACC) has been shown to be involved in several aspects of prediction and risk processing, including error detection (Falkenstein et al, 2000; Gehring et al, 1993) conflict monitoring (Botvinick et al, 2001; Botvinick et al, 2004) and, more recently, the generation of prediction signals (the PRO model;

Alexander & Brown, 2011). The PRO model, in particular, hypothesizes that the mPFC is primarily involved in simulating different potential outcomes for an executed action. This is in contrast to model-free, or habit-based learning, in which expected values are cached and associated with specific action-outcome associations; as a result, they are much more computationally efficient, but are relatively inflexible as opposed to model-based reinforcement learning (Daw et al, 2006). For addict populations, it is plausible that for action-outcome associations for the self-administration of drugs represents an extreme bias toward a model-free reinforcement learning paradigm, which is highly inflexible even when the expected value associated with these actions becomes significantly devalued.

Expected Value Mediated by Prediction Values

In light of these findings, studies have been carried out in order to uncover the relationship of prediction signals to the expected value of receiving a drug, and what cortical mechanisms modulate the processing of these prediction signals. To this end, mediation analyses have been carried out in order to delineate how interactions between cortical activity and behavioral measures might contribute to addiction-related behavior. For example, in a study by Krawitz et al (2011), participants with schizophrenia were compared to controls during a change-signal and delayed matching-to-sample task (DMTS). Although individuals with schizophrenia and drug addiction populations may appear to be separate categories of mental disorder, it is possible that in both cases a dysfunctional prediction system underlies the overt pathology. Relative to controls, participants with schizophrenia exhibited a reduced activation profile within the perigenual ACC in response to cues predicting error likelihood, suggesting that hypoactivation in response to violations of expected outcomes was being driven by a dysfunctional neural response

to the probability of receiving an error. As working memory was not significantly different between controls and individuals with schizophrenia, it was unlikely that the difference in error likelihood and error unexpectedness is driven by working memory differences.

Therefore, it is plausible that a similar paradigm employing the same mediation analysis reported in the Krawitz et al (2011) study would reveal a similar pattern of lower levels of perigenual ACC activation in response to error unexpectedness being driven by lower levels of error likelihood effects in drug addiction populations. Future neuroimaging studies could take this analysis further by examining the changes in connectivity strength between distinct regions of medial PFC, with higher levels of rostral and perigenual ACC activation in response to error likelihood cues leading to heightened connectivity with outcome-related areas of dorsal and rostral caudal zone areas of the anterior cingulate.

Reward Prediction Errors and State Prediction Errors

A closely related theoretical paradigm of action-outcome prediction focuses on whether behavior is a more habitual process which relies on ingrained action-outcome associations, or whether the behavior of an organism can be better modeled by deliberative processes in which predictions are formulated about the potential outcomes of each possible action before selecting among them. As an example of applying this model-based process to human subjects, Glascher et al (2010) examined the neural underpinnings of these two types of learning: 1) Model-free, and 2) Model-based, or forward models. Model-free learning is updated through reward prediction errors (RPEs), reflecting a discrepancy between the actual and expected reward. Action-value associations are thus learned through this process. Model-based reinforcement learning, on the other hand, involves the simulations of potential action-outcome associations based on prior

experience or educated guesses, and the probability and magnitude of reward or punishment associated with those actions. This consists of the construction and evaluation of state prediction errors (SPEs), which compare the current environment one is in against a previous environment, and computes any differences in expected reward between the two, which requires conscious deliberation between the alternatives.

In the experiment by Glascher and colleagues (2010), these models were applied to a reinforcement learning task in which subjects had to learn the values associated with fractals, which presumably would not contain any a priori information or emotional valence. In the first session, participants were exposed to the different possible decision-making trees without making a response, which provided a pure estimate of SPE. Estimates of model fit were derived from neural activity associated with these state prediction errors. Significant activity was found in bilateral IPS and DLPFC for state prediction errors, while ventral striatal activity was found only for RPEs. Interestingly, only intraparietal sulcus (IPS) activity was observed for the first scanning session, since this region appears to be involved in spatial memory; however, once the various states and contingencies are learned by exploring the tree of possible outcomes, the need to encode this information is reduced, as was shown through attenuated BOLD signal within the IPS for the remainder of the experiment.

A similar model-based approach could reveal important differences between drug addiction populations and controls when examining responses to RPEs and SPEs. Presumably, both RPEs and SPEs are not calculated optimally in drug addiction populations, which accounts for their observed maladaptive decision-making during tasks such as the Iowa Gambling Task. Associated cortical and subcortical areas contributing to the computation of RPEs and SPEs would, therefore, be assumed to be deficient or malfunctioning in drug addiction populations.

Currently, a study is being conducted involving the direct application of nicotine vapor to subjects while undergoing fMRI scanning. This study was preceded by a behavioral study validating the use of the nicotine delivery device by measuring cotinine, a nicotine metabolite, in the blood as subjects inhaled nicotine vapor through the device. The results of this study showed that the amount of nicotine in the blood could be estimated from the amount of vapor inhaled through the device, in addition to individual measures such as weight (de Mendizabal et al, 2014). Such an approach makes an fMRI approach more feasible, as the estimated amount of nicotine in the blood could be used as a covariate when examining neural responses to decisions involving drugs (see Appendix C).

1.4 Rationale for the Present Studies and Hypotheses

Three studies are discussed in the following chapters, which as a whole compare different models of mPFC function and expand upon critical aspects of the PRO model. In the first study, a comparison of the conflict monitoring theory and the PRO model, examines whether merely imagining an error reveals similar patterns of activity to actually committing an error. Action values are thought to be represented in part in the dorsal and ventral medial prefrontal cortex, yet current studies have focused on the value of executed actions rather than the anticipated value of a planned action. Thus, little is known about the neural basis of how individuals think (or fail to think) about their actions and the potential consequences before they act. We scanned individuals with fMRI while they thought about performing actions that they knew would likely be correct or incorrect. In this study we show that merely imagining an error, as opposed to imagining a correct outcome, increases activity in the dorsal anterior cingulate cortex, independently of subsequent actions. This activity overlaps with regions that respond to actual

error commission, revealing a distinct network that signals the prospective outcomes of one's planned actions. As this specific contrast of imagined versus actual errors occurs in the absence of any overt motor conflict, the conflict monitoring model is unable to adequately explain these findings.

The second study discussed is a direct application of the PRO model in generating regressors and applying it to fMRI data (Jahn et al, 2014). A number of theories have been proposed to account for the role of anterior cingulate cortex (ACC) and the broader medial prefrontal cortex (mPFC) in cognition. The recent Prediction of Response Outcome (PRO) computational model casts the mPFC in part as performing two theoretically distinct functions: learning to predict the various possible outcomes of actions, and then evaluating those predictions against the actual outcomes. Simulations have shown that this new model can account for an unprecedented range of known mPFC effects, but the central theory of distinct prediction and evaluation mechanisms within ACC remains untested. This study uses combined computational neural modeling and fMRI to demonstrate that prediction and evaluation signals are indeed each represented in the ACC, and furthermore, they are represented in distinct regions within ACC. This study achieves this by independently manipulating both the number of predicted outcomes and the degree to which outcomes violated expectancies, the former providing assessment of regions sensitive to prediction and the latter providing assessment of regions sensitive to evaluation. Quantitative regressors derived from the PRO computational model show that prediction-based model signals load on a network including the posterior and perigenual ACC, but outcome evaluation model signals load on the mid-dorsal ACC. These findings are consistent with distinct prediction and evaluation signals as posited by the PRO model and provide new perspective on a large set of known effects within ACC.

The third study is a summation of several different strands of modeling work and empirical studies performed on the mPFC. First, given that other neuroimaging studies have found mPFC activity in response to the unexpected absence of pain (Garofalo et al, 2014) and violations of predictions based on cues for upcoming Stroop stimuli (Aarts et al, 2011), this study sought to combine prediction error across both painful stimuli and Stroop stimuli contexts into a single factorial design. The purpose was to test whether the mPFC signals prediction errors in a single homogeneous or several heterogeneous regions, and whether these prediction error signals are processed similarly regardless of valence (Ferdinand et al, 2014). Furthermore, this design would provide a direct test against other competing models of mPFC function, most notably the reinforcement learning model (Holroyd & Coles, 2002).

Chapter 2

The neural basis of predicting the outcomes of planned actions

2.1. Introduction

A key feature of human intelligence is the ability to predict the outcomes of one's own actions prior to executing them. Much of the literature on decision-making and reinforcement learning focuses on learning the value of various available options. The optimal decision is one that has the highest value in the decision-maker's subjective evaluation (Thorndike, 1911), with perhaps some value on exploring new options (Kaelbling et al., 1996). Environmental cues indicate what options are available, and the cues in turn guide instrumental responding via learned stimulus-response (S-R) associations (Sutton and Barto, 1998). This is the essence of model-free reinforcement learning (Dayan and Niv, 2008). Such constitutes an *inverse model* (Shadmehr and Wise, 2004), in that stimulus cues (S) activate a representation of the desired goal such as a piece of food, and this goal is mapped backward to the response (R) necessary to achieve the goal. The values of stimuli and the goals they represent are likely represented in the orbitofrontal cortex (OFC) (Tremblay and Schultz, 1999; Schoenbaum et al., 2003). All of this works fine for habit learning.

The situation is more difficult when an animal faces a novel environment in which the S-R association has not been learned, or there is a more complex set of constraints, so that there is no one automatic best response. This is where *forward models* as in model-based reinforcement learning (Shadmehr and Wise, 2004; Daw et al., 2005; Glascher et al., 2010) are useful. A forward model predicts the outcome of a planned action. This is a learned response-outcome (R-O) association (Colwill and Rescorla, 1990) which affords a "dynamic evaluation lookahead"

(van der Meer and Redish, 2010). Favorable outcome predictions might further activate the corresponding response plan, while unfavorable or risky outcome predictions might suppress it.

The process of employing a forward model to predict the likely outcomes of planned actions is akin to the popular notion of thinking before acting. Humans can think about or imagine (with varying accuracies) what might be the outcome of a planned action. Nonetheless, relatively little research has been done on the neural basis of thinking ahead, with just a few cognitive (Johnson, 2000; Hassabis et al., 2007), neuroimaging (Newman et al., 2009; Glascher et al., 2010), and rat (van der Meer and Redish, 2010) studies. Some results suggest that anterior cingulate cortex (ACC) may be involved in anticipating adjustments in control (Sohn et al., 2007; Aarts et al., 2008; Aarts and Roelofs, 2011). We previously showed that the medial prefrontal cortex (mPFC), and especially ACC may learn to predict the likelihood of an impending error resulting from current actions (Brown and Braver, 2005). Here we use fMRI to ask whether and how the ACC may signal the error likelihood of imagined responses, as distinct from the alternative hypothesis that ACC is activated only by impending actions. We use a simple task that isolates the outcome prediction by asking subjects to imagine performing an action and experiencing its consequences, while controlling for the subsequent action execution.

2.2. Methods

Participants

Data from 22 right-handed participants were collected (mean age = 23.42, SD = 2.80). Data from two participants were discarded due to insufficient correct responses and data from one participant was excluded due to a scanning artifact, leaving 19 usable participants (11 female). Participants reported no history of psychiatric or neurological disorder, and reported no current

use of psychoactive medications. Participants were compensated $25/hour for their time. Participants were trained on the task on a computer outside of the scanner until they gave verbal confirmation that they understood the task. The experimenter observed the participant's performance and judged whether they demonstrated sufficient understanding of the task.

Participants were informed that they would receive compensation based on their performance, although they were unaware of how much they would receive for rewarding feedback. In reality, they received $0.05 for each rewarded outcome (described in further detail below in section "Experimental Paradigm").

Experimental paradigm

The task consisted of two phases: an imagine phase and a response phase. During the imagine phase, participants were instructed to imagine the consequence of making particular responses. During the response phase, participants were instructed to choose one of two possible responses. The appropriate response was determined by feedback history. When a particular response was rewarded, participants were instructed to make that response again. If a response was not rewarded, participants were instructed to make the alternate response. Hence, prior to each trial, participants had a belief about the correct response and could therefore imagine the consequences of a response that matched that belief (i.e. imagine correct) or violated it (i.e. imagine error).

On each trial, the imagine phase began with the sequential presentation of two white arrow cues on a black background, with one pointing left and the other pointing right (Figure 1). The order of presentation of the arrow cues was counterbalanced. Participants were instructed to simply imagine themselves pressing the corresponding left or right buttons with the left or right

index finger, along with the corresponding outcome they would expect if they were to actually press the button. Both left and right responses were imagined separately on each trial, so the probabilities of imagining each event were equal. After a variable delay, the response phase was signaled by an exclamation mark ("!") which cued them to respond with either a left or right actual button press. Crucially, the responses that they imagined were independent of the actual response that they made. Participants would then be presented with a "$" or a "0" as feedback. A "$" would mean that they had gained a point, while a "0" would mean that they had gained nothing. Participants were informed that if they were rewarded on a trial (i.e., if they received a "$" as feedback) that they should make the same button response on the next trial, but a "0" indicated that they should switch. The correct button (left vs. right) switched across trials with a relatively low probability. With this design, subjects could predict the outcome of each possible button press with moderate confidence. This allowed us to examine neural activity related to imagining distinct error and correct responses without confounding the results with a particular effector. The probability of an underlying switch was 0 for the first two trials following a switch, then 0.33 per trial for trials three through seven, and 1.0 after eight trials. This distribution ensured that switches occurred but were unpredictable and less likely than chance. After receiving feedback, participants were presented with a blank screen that lasted either 1, 3, 5, or 7 seconds, based on an exponential distribution function (Dale, 1999).

On 20% of the trials, a question mark ("?") was presented instead of arrow cues. During this condition, participants were instructed to recall the last response they had made and the corresponding outcome they had received, whether it was an outcome signaling a reward or not gaining a reward. When the exclamation mark cue was presented, participants were to make the

same response they had made on the previous trial, whether it was rewarded or not. These trials were included for purposes not relevant here and were modeled separately.

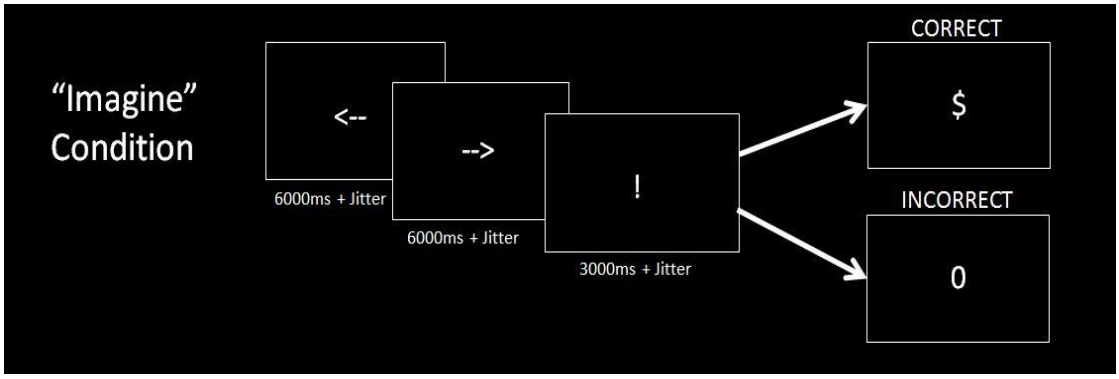

Figure 1: Imagine condition. In the Imagine condition, participants saw a sequence of two arrows, one facing left and the other facing right (order randomized across trials). As each arrow appeared, participants were instructed to imagine performing the corresponding button press response (left or right) and the outcome associated with it. An exclamation mark ("!") cued the subjects to make a response of their choice. One of the two options was correct, and the other would be incorrect. The correct response in the preceding trial was more likely to be correct in the current trial. Participants received either rewarded ("$") or non-rewarded ("0") feedback as a result of their choice. The response cue and outcome cues were identical to the Imagine condition.

fMRI acquisition and data preprocessing

The experiment was conducted with a 3 Tesla Siemens TIM Trio scanner using a 32-channel head coil. Foam padding was inserted around the sides of the head to increase participant comfort and reduce head motion. Imaging data was acquired at a 30° angle from the anterior commissure-posterior commissure line in order to maximize signal-to-noise ratio in the orbital and ventral regions of the brain (Deichmann et al., 2003). Functional T2* weighted images were acquired using a gradient echo planar imaging sequence [30 x 3.8mm interleaved slices; TE = 25ms; TR = 2000ms; 64x64 voxel matrix; 220x220mm field of view]. Three runs of data were

collected with 240 functional scans each. High resolution T1-weighted images for anatomical data [256x256 voxel matrix] were collected at the end of each session.

SPM5 (Wellcome Department of Imaging Neuroscience, London, UK; www.fil.ion.ucl.ac.uk/spm) was used for preprocessing and data analysis. The functional data for each run for each participant was slice-time corrected and realigned to each run's mean functional image using a 6 degree-of-freedom rigid body spatial transformation. The resulting images were then coregistered to the participant's structural image. The structural image was normalized to standard Montreal Neurological Institute (MNI) space and the warps were applied to the functional images. The functional images were then spatially smoothed using an 8mm Gaussian kernel.

fMRI analysis

Functional neuroimaging data were analyzed using a general linear model (GLM) with random effects. Feedback for correct and incorrect responses were modeled with a canonical Hemodynamic Response Function (HRF) at the time of feedback. Two regressors modeled each imagine event. A delta regressor locked to the onset of stimulus presentation was included to capture initial perceptual activation. An epoch regressor onsetting 1 second after stimulus presentation and spanning the duration of the imagine event was included to capture the act of imagining itself. These epoch regressors are the regressors of interest for present purposes. Separate regressors were included for imagine error and imagine correct events.

Additional regressors modeled left vs. right button presses. Contrasts were conducted on imagining a potential error outcome (ImagineError) compared to imagining a potential correct outcome (ImagineCorrect). This contrast would reveal whether there was significantly more

activity for merely imagining an error outcome as opposed to a correct outcome. Separate contrasts were computed for each subject, and results are based on a group-level random effects analysis on these contrasts.

Unless otherwise stated, all whole-brain results were thresholded at $P < 0.001$ at the voxel-level with a 121 voxel cluster extent providing a corrected $p < 0.05$ threshold according to AlphaSim. In order to interrogate whether ACC regions involved in error feedback processing are also involved in imagining an error, an additional analysis for the ImagineError – ImagineCorrect contrast was assessed using a functional mask from the FeedbackError – FeedbackCorrect contrast. Results within this small volume were thresholded at $p < 0.05$ at the voxel-level with a 174 voxel cluster extent providing a corrected $p < 0.05$ threshold according to AlphaSim.

2.3. Results

Behavioral Results

Behavioral data were analyzed in order to confirm that subjects performed the task appropriately. If participants successfully followed instructions on either switching or repeating their response on the next trial, participants would on average receive 17 reward outcomes per run, or 51 reward outcomes over all three runs. On average, participants performed the task at a satisfactory level (mean number of reward outcomes per run = 15.95, SD = 1.02). Participants who received 12 or fewer reward outcomes for two or more runs were excluded from further analysis.

A subset of participants (N=10) were given a debriefing survey after scanning asking whether they were able to visualize the motor response associated with each arrow, whether they were able to imagine the outcome associated with each button press, and whether they felt

motivated to respond to gain the bonus money. Ratings were made on a Likert scale from 1 to 5, with 1 being the lowest confidence in the given response and 5 being the highest. In general, participants rated that they were able to visualize the motor response (mean rating = 4.3) and able to imagine the outcome associated with each button press (mean rating = 4.7). Participants also appeared to be motivated to perform the task well (mean rating = 4.6). A Wilcoxon Signed-Rank test showed that all ratings were significantly different from an average score of 3, which would represent indifference toward each of the questions (all P's < 0.01). Hence, the behavioral data indicated that subjects understood and performed the task as instructed.

Imaging Results

We began by confirming that error feedback produced heightened activation in the ACC compared to correct feedback as would be expected by prior literature (Hohnsbein et al., 1989; Gehring et al., 1990). Confirming these activations, the contrast of FeedbackError – FeedbackCorrect produced robust activations in the dorsal ACC and pre-SMA, as well as lateral frontal and parietal regions. These results indicate that the paradigm appropriately elicited error signals in the ACC.

Next, we examined the neural correlates of imagining erroneous actions. A whole-brain contrast of ImagineError – ImagineCorrect did not reveal any significant clusters at our strict corrected threshold. However, at a more liberal threshold ($p < 0.005$ uncorrected), there was some evidence of heightened activation in the ACC, but no other frontal region. Based on our *a priori* hypothesis about the role of the ACC in error prediction, we looked for evidence of increased ACC activation within a small volume defined by the FeedbackError – FeedbackCorrect contrast (see Methods). Within this ROI, a significant effect of ImagineError-

ImagineCorrect was found in the dorsal region of the ACC and pre-supplementary motor area (pre-SMA) (Figure 2; MNI -10, 10, 46; k = 217 voxels; peak voxel z-value = 3.18, $p < 0.05$, cluster-corrected). We have proposed that ACC activity signals in part the likelihood of an error in a particular condition (Brown and Braver, 2005; Brown and Braver, 2007), as part of a more general function of predicting the outcome of an action (Alexander and Brown, 2010). Our finding here of greater activity for imagining errors relative to imagining correct outcomes is consistent with this possibility. This leads to a pair of follow-up questions, namely (1) which parts of the brain might drive the apparent prediction signal in the ACC, and (2) what might account for the greater activity when imagining errors relative to imagining correct responses? In answer to the first question, one possibility is that outcome predictions are driven by motor-related activation representing the imagined plan to move, as seen in the lateralized readiness potential (LRP)(Kornhuber and Deecke, 1965). In that case, activating a plan to move in the motor cortex might in turn activate the ACC to represent the movement plan as well as its anticipated outcome. We previously showed that activating a greater number of movement plans (even without response conflict) could lead to greater ACC activity (Brown, 2009), but it was unclear whether activating the plans alone without the corresponding execution would be sufficient to activate the ACC. This was a key unresolved question, as we have hypothesized elsewhere that ACC activity represents the prediction of an action outcome, which can be made in advance (or perhaps even independently) of action execution (Alexander and Brown, 2010).

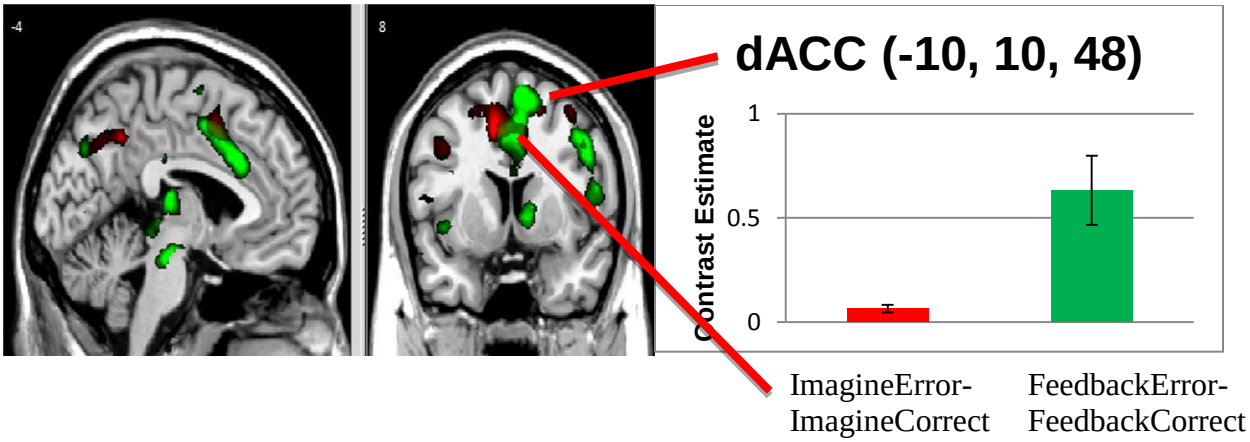

Figure 2: Red: Region of the dorsal ACC showing increased activation in response to imagining an error contrasted with imagining a correct response, shown at $p < 0.05$, uncorrected. Green: Overlapping area for receiving nonrewarding feedback contrasted with receiving rewarding feedback, shown at $p < 0.001$, uncorrected.

In answer to the second question of what accounts for greater ACC activity for imagining errors than correct responses, one possible account is that even when subjects imagine an error, they also maintain an active representation of the correct response as they subsequently intend to execute it. This would lead to greater summed motor cortex activity when imagining errors relative to imagining correct responses, which in turn would lead to greater ACC activity under both our outcome prediction model of the ACC (Alexander and Brown, 2010) and the conflict monitoring model of the ACC (Botvinick et al., 2001), although as we have shown, such effects do not depend on response conflict *per se* (Brown, 2009).

To address these questions, we identified regions in motor cortex (Areas 4 and 6) that showed effects of executing particular responses, i.e. RespondLeft > RespondRight (right motor cortex, MNI 46, -28, 54, k = 2161 voxels) and RespondRight > RespondLeft (left motor cortex, MNI -34, -32, 46, k = 3437 voxels) at a cluster corrected threshold of $P < 0.001$. We then tested whether the left motor cortex was more active when imagining a left response that was an error vs. imagining a left response that was correct, which would indicate greater activation for the

alternative correct response when imagining an error, relative to the alternative error response when imagining a correct response. The results were consistent with this hypothesis. The results of the contrast in the left motor cortex were significant (Imagine/Left/Error > Imagine/Left/Correct, MNI -44, -28, 60; k = 1563 voxels; peak z-value = 5.59; $P < 0.001$, cluster corrected). A similar result held for the right motor cortex (Imagine/Right/Error > Imagine/Right/Correct, MNI 38, -18, 48; k = 429 voxels; peak z-value = 5.15; $P < 0.001$, cluster corrected). These results suggest that the greater activity in ACC for imagining an error vs. imagining a correct response may derive from the combined signal of activities in motor cortex that reflect planned responses (Brown, 2009), regardless of whether or not those responses are not subsequently executed.

Chapter 3

Distinct regions of anterior cingulate cortex signal prediction and outcome evaluation

3.1. Introduction

The study just discussed was one of the first empirical fMRI tests of the PRO model against a competing theory of mPFC function. A natural follow-up to the previous study was to apply the PRO model to a task where prediction error was modulated on a trial-by-trial basis, with certain trials carrying greater prediction error than others on the basis of how often a given outcome occurred on that trial. While the PRO model provides a compelling unified theory of ACC function, its main proposal remains untested, namely whether distinct prediction-related and outcome-related signals exist within the ACC. If so, a related question is whether the distinct prediction and outcome signals are found in overlapping regions of ACC, or whether they are largely segregated within different subregions of ACC. The PRO model predicts only that the two signals will exist; it does not predict whether or not they will overlap within regions of the ACC. The current evidence of regional distinctions within the ACC suggests that these two signals may not only exist within ACC but also be spatially distinct. For example, several recent studies have outlined distinct subregions of the ACC based on probabilistic connectivity (Beckmann et al, 2009), dynamic causal modeling (Fan et al, 2008), motor representations (Amiez and Petrides, 2012), neural deficits in schizophrenia (Krawitz et al., 2011), and experimental paradigms incorporating error, conflict, and task-switching effects into a single design (Nee et al., 2011), highlighting the anatomical and functional heterogeneity of the ACC. The current study was designed to test whether a model-based analysis could identify these prediction and outcome processes in the ACC, and if so, whether these processes are spatially distinct or overlapping.

Here we find that distinct regions of the ACC are involved in generating prediction and outcome (i.e. prediction error) signals, in line with the PRO model. To investigate this, we use a task which parametrically manipulates both the number of predictions subjects make and the number of surprising outcomes. We present the same behavioral sequence to both the PRO model and human subjects, and we derive model-based regressors from the PRO model. These are entered as covariates in the fMRI analysis to identify regions that correspond to the theoretical components of the PRO model.

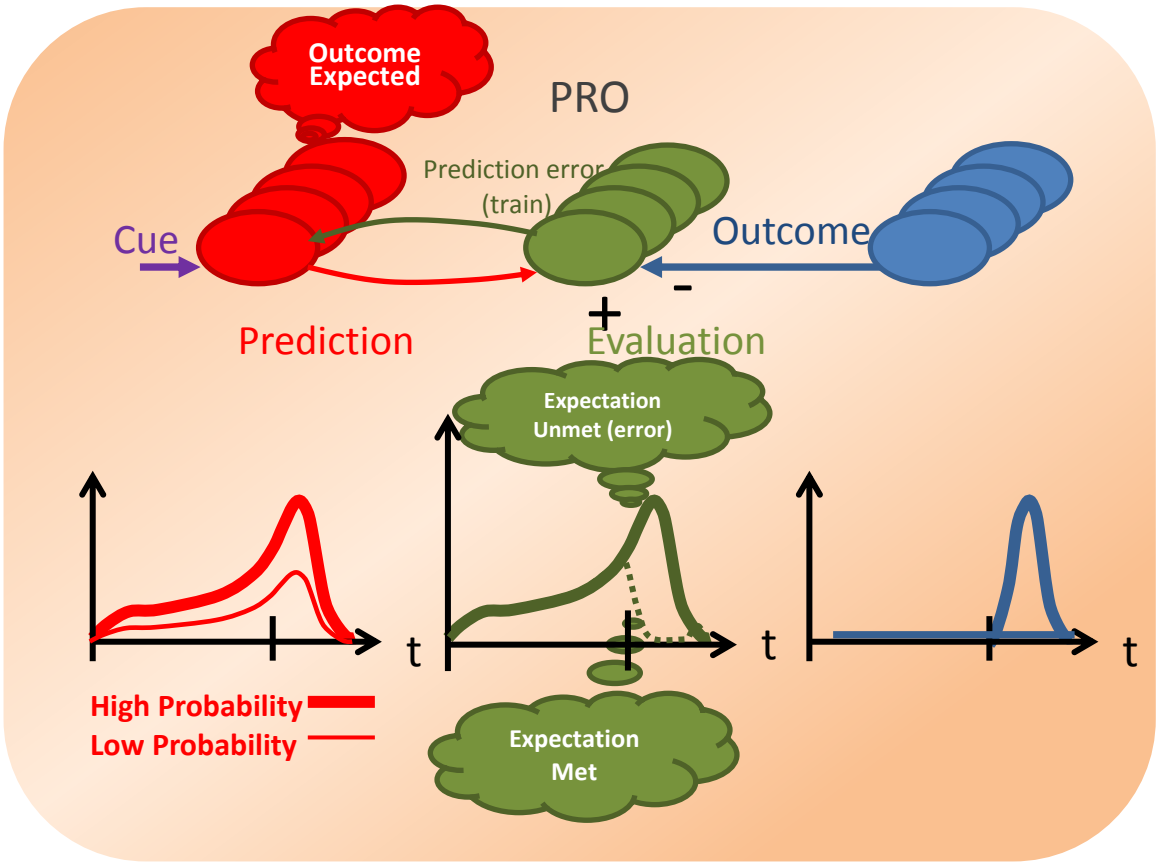

Figure 1: A general conceptual illustration of the Predicted Response Outcome (PRO) model of anterior cingulate cortex. Left: The Prediction units (red) generate a timed prediction of what outcomes are expected, with what probability, and when. There are four Prediction units (ellipses) shown here for illustration purposes, corresponding to four predicted outcomes, although the number of Prediction units will vary in general with different tasks so that there is one Prediction unit for each corresponding possible outcome. Lower left: Greater probabilities (y-axis) are

associated with greater prediction activity, which peaks at the time when the outcome is expected due to the temporal discounting of the probability. Middle: The evaluation units compute *negative surprise*, i.e. they detect when an expected outcome fails to occur. This is simply the difference between the predicted and actual outcomes and represents how improbable the negative surprise was. The green arc from Evaluation to Prediction indicates that prediction errors from the Evaluation units train and update the prediction signals. Lower middle: Events that are predicted with a high probability yield greater surprise signals (Expectation Unmet) when they fail to occur, but a weaker surprise signal when they do occur (Expectation met). Right: Outcomes occur (i.e. Outcome value rises rapidly and transiently to 1) or fail to occur (i.e. Outcome value is 0) at specific times.

3.2. Materials and Methods

The Institutional Review Board of Indiana University approved the experimental procedures reported here.

Participants

Data from 14 right-handed participants (9 female) were collected (mean age = 24.93, SD = 2.92). Participants reported no history of psychiatric or neurological disorder, and reported no current use of psychoactive medications. Participants were compensated $25/hour for their time, in addition to a performance bonus based on how many correct responses they made during the task. Participants were trained on the task on a computer outside of the scanner until they gave verbal confirmation that they understood the task.

fMRI Paradigm

The task was designed to manipulate the neural activity related to predicting and evaluating outcomes. To achieve this, we manipulated the number of outcomes subjects had to predict as

consequences of their actions, as well as the degree to which the actual outcomes differed from the predicted outcomes. Subjects were instructed to make two choices regarding a pair of options, and then they were required to predict the outcomes of the choices. Critically, for some trials, participants were required to maintain predictions about each outcome from their pair of choices (*Predict2* condition), while on other trials, participants were required to maintain a prediction about only one outcome from their pair of choices (*Predict1* condition). These conditions were later contrasted to test for an effect of an increasing number of maintained predictions. Thereafter, the subjects were informed of the outcomes of their choices. Outcomes could violate zero, one, or two predictions thereby providing a parametric effect of expectancy violation. Each participant underwent a behavioral session outside of the scanner consisting of 100 trials. If the participant felt that they understood the task and consented to undergo the scanning paradigm at a later time, each participant completed another 50 trials immediately before scanning to refresh their memory of the task. During scanning each participant underwent 5 runs of 100 trials each, with each run lasting 8 minutes and 40 seconds.

The task consisted of three phases: A choice phase, a prediction phase, and an outcome phase (See Figure 2). During the choice phase, participants were presented with two rows of two boxes, forming a pair of boxes for each row. The two rows of boxes were separated by a white horizontal line. A question mark ("?") placed in between a row of boxes prompted subjects to choose one box from the row. Choices were to be based on prior outcomes (described below). After choosing between boxes in one row, the question mark moved to the other row. The placement of the first prompt (top or bottom) was randomly counterbalanced across trials.

During the choice phase, participants chose one box from each row. In the *Predict2* condition, all boxes contained a question mark ("?") which informed the subject that a chosen

box would yield an outcome cue (described below). In the *Predict1* condition, one of the rows contained boxes with "X"s, which indicated that the chosen box would not yield an outcome. The other row of boxes contained question marks and thus yielded outcomes. As a result, in the *Predict2* condition, subjects made two choices that yielded outcomes, while in the *Predict1* condition, subjects made two choices, only one of which yielded an outcome. Outcome cues informed subjects whether to choose that row's box again on the next trial (i.e., a "stay" cue), or to choose the other box in that row on the next trial (i.e., a "switch" cue). Hence, the results of outcomes had to be maintained in order to inform future choices. After the participants made their choices, they were presented with their choices for 1000ms, followed by a fixation cross of a jittered duration, before beginning the prediction phase.

During the prediction phase, participants were re-presented with their choices. This phase signaled to the subject that outcomes would soon be presented and provided a cue for the prediction of those outcomes. Hence, the prediction phase was used to model prediction. Although prediction of outcomes may begin immediately after choices are made, because of the prediction phase's closer temporal proximity to the actual outcome received by the participants, the PRO model expects prediction-related cells to ramp up during the prediction phase (Alexander & Brown, 2011), a phenomenon that has also been observed empirically (Amador et al., 2000; Hayden et al., 2009; Shidara and Richmond, 2002). Furthermore, this was designated as the prediction phase because it was dissociated from the motor activity preceding it. In the *Predict2* condition, both boxes had question marks in the center, signaling that there would be two outcome cues, and that the participant should maintain two separate outcome predictions. In the *Predict1* condition, only one box had a question mark in the center while the other box had an "X" in the center, signaling that there would be only one outcome cue and that the participant

should therefore maintain one outcome prediction. The prediction phase lasted for a jittered duration up to 7500ms, and was followed by the outcome phase. The condition of predicting two outcomes instead of one outcome was not confounded with working memory load, because even in the *Predict1* condition, subjects had to remember the location of the unpredicted outcome in order to choose it correctly in subsequent trials.

Outcomes of the participants' choices were revealed in the outcome phase. In the *Predict2* condition, both of the chosen boxes revealed an outcome cue. In the *Predict1* condition, only one of the chosen boxes revealed an outcome cue. Subjects were instructed that in the *Predict1* condition, the box with the "X" in the center would not reveal an outcome cue. Subjects were instructed that the outcome of their choices would inform what decision to make on the next trial. Outcomes informed subjects whether to choose the same box as the current trial (stay) or choose the other box (switch). In this way, subjects were motivated to attend to the feedback and update their choices accordingly in subsequent trials. Stay and Switch cues were denoted by "*" and "0" with stay/switch to character mappings counter-balanced across subjects. If subjects performed the task correctly, they would expect to find a stay cue most of the time (p=0.6) in all chosen boxes, and in at least one box if two outcomes were predicted (p=0.8). As a result, we expected that participants would predict a stay cue and that a switch cue would be a violation of that prediction.

Subjects were told that they would receive a reward of $0.05 on every trial if they correctly followed the outcome cue (either switching their response or making the same response on the next trial), and that they would not receive any reward if they failed to follow the outcome cue. This dissociation of receiving either a switch or stay cue and earning reward ensured that the observed effects were not confounded with reward anticipation or error likelihood

Feedback provided information for what options to choose on the next trial. Of particular interest was the phase following the choice phase. During this prediction phase, subjects maintained predictions about the outcome of their choice(s), affording the assessment of prediction-related neural activation. The task was designed to separate these phases of prediction and evaluation so that BOLD responses to each could be estimated independently.

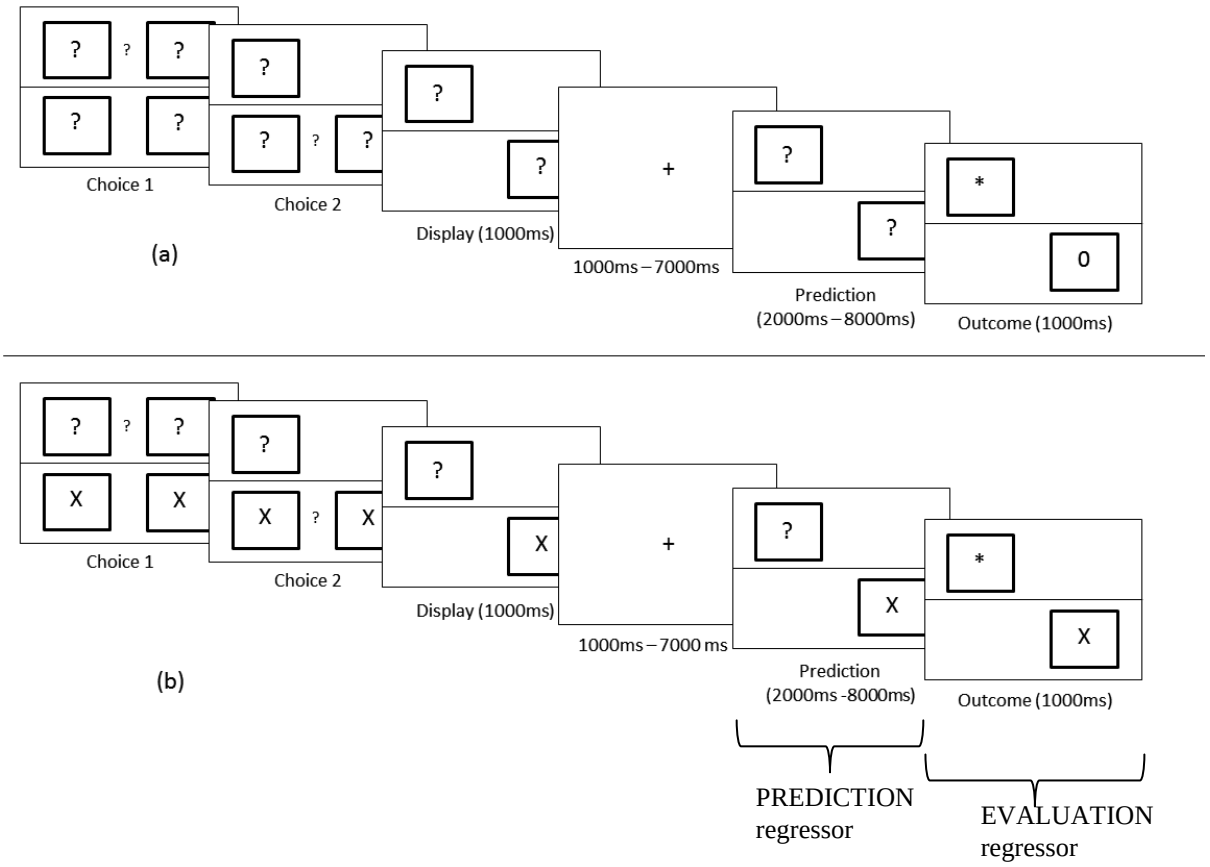

Figure 2: Task design. Dissociation of decision, prediction, and outcome effects. In each trial, the question mark between the two upper or two lower boxes prompted a choice between the adjacent boxes. After a delay to separate out the decision and motor-related activities, subjects were re-presented with their choices, at which point they could predict the impending feedback to be delivered in the Outcome phase of the trial. Outcomes ("*" and "0") shown later in the chosen boxes indicated that subjects should choose the same box again on the next trial (stay) or, less commonly, choose a different box on the next trial (switch). (a) In the *Predict2* condition, participants made a series of two choices and later received outcome feedback for each choice. During the prediction phase of the task,

participants could predict the impending outcome(s). (b) In the *Predict1* condition, participants also made two choices, but they knew that they would receive only a single outcome cue. No feedback was given for boxes containing an "X". Instead, participants were told to repeat their previous choice in the next trial. Hence, there was only a single outcome to predict and evaluate in the *Predict1* condition. Overall, the task affords distinct activity estimates of decisions, of predicting 1 vs. 2 outcomes, and of receiving 0, 1, or 2 rare switch feedback cues. The actual stimuli were shown to subjects with the colors inverted, i.e. white stimuli on black background.

Control paradigm

Thirteen (13) of the participants also underwent a control run before performing the experimental task (one participant failed to complete the control run due to technical issues). During this run, participants were presented with the different combinations of box locations as they would see during the Prediction phase of the experiment (see Figure 2). These were the only stimuli presented during the control block. The presentations of the boxes were separated by a jittered interval with a similar distribution to the jitters during the actual experiment. Participants were instructed to attend to the boxes as they normally would during the experiment and to remain focused. In contrast with the experimental condition described above, they were not otherwise required to respond, and participants were instructed not to form predictions as they would in the experimental task because the computer automatically made choices for the subjects.

The purpose of this control run was to determine whether brain activation for the contrast of predicting two outcomes vs. predicting one outcome (see Imaging Results below) could be explained by visual, eye movement, or attention factors instead of monitoring for multiple action outcomes.

3.2.1. fMRI analysis

Image acquisition and preprocessing

The experiment was conducted with a 3 Tesla Siemens Trio scanner using a 32-channel head coil. Foam padding was inserted around the sides of the head to increase participant comfort and reduce head motion. Imaging data was acquired at a 30° angle from the anterior commissure-posterior commissure line in order to maximize signal-to-noise ratio in the orbital and ventral regions of the brain (Deichmann et al., 2003). Functional T2* weighted images were acquired using a gradient echo planar imaging sequence [30 x 3.8mm interleaved slices; TE = 25ms; TR = 2000ms; 64x64 voxel matrix; 220x220mm field of view]. For the experimental condition, five runs of data were collected with 240 functional scans each. For the control condition, one run of data was collected with 145 functional scans. High resolution T1-weighted images for anatomical data [256x256 voxel matrix] were collected at the end of each session.

SPM5 (Wellcome Department of Imaging Neuroscience, London, UK; www.fil.ion.ucl.ac.uk/spm) was used for preprocessing and data analysis. The functional data for each run for each participant was slice-time corrected and realigned to each run's mean functional image using a 6 degree-of-freedom rigid body spatial transformation. The resulting images were then coregistered to the participant's structural image. The structural image was normalized to standard Montreal Neurological Institute (MNI) space and the warps were applied to the functional images. The functional images were then spatially smoothed using an 8mm Gaussian kernel.

Model-Based Analysis

The PRO model characterizes dACC/mPFC as a region involved with learning to predict likely outcomes and signaling unexpected deviations from predicted outcomes. The model learns temporally discounted estimates of the likelihood of possible outcomes using a temporal difference (TD) learning algorithm (Sutton & Barto, 1990) that has been extended in the following ways. First, the PRO model learns predictions for multiple, independent outcomes, regardless of their affective valence, in contrast to TD learning which learns the aggregate reward value of outcomes weighted by the frequency with which those outcomes are observed. Second, the PRO model generates a vector-valued error signal in order to update model predictions regarding likely outcomes according to the following equation:

$$\delta_{i,t} = O_{i,t+1} + \gamma P_{i,t+1} - P_{i,t} \quad (1)$$

where O is a vector reflecting the occurrence of outcomes *i* at time *t+1*, P reflects outcome predictions, and γ is a discount factor ($\gamma = 0.95$). Model predictions were computed as

$$P_{i,t} = \sum_j I_{j,t} W_{i,j,t} \quad (2)$$

where *I* is a vector of binary values reflecting the presence (1) or absence (0) of a particular input *j* at time *t*, and W is the matrix of weights indicating the discounted estimate of the likelihood of an outcome *i* for all inputs. Model weights are updated according to

$$W_{i,j,t} = W_{i,j,t} + \alpha \delta_{i,t} \bar{I}_{j,t} \quad (3)$$

where α is a learning rate parameter ($\alpha = 0.1$) and $\bar{I}$ is an eligibility trace computed as

$$\bar{I}_{j,t} = I_{j,t} + 0.95 \bar{I}_{j,t} \quad (4)$$

. In previously published simulations, the error signal δ was used to dynamically adjust the rate at which new information (in the form of unexpected deviations from expectations) was integrated into the model to allocate top-down control of behavior, allowing the PRO model to fit

observed, aggregate behavioral data. For our current analysis, our aim is different in that, rather than fitting behavioral data, we seek to generate trial-by-trial predictions of ACC activity for individual subjects using the sequence of outcomes observed by those subjects in the course of the task. Accordingly, we simulate the PRO model during the period in each trial following the presentation of the predict cue and terminating following feedback. There were two model inputs used during simulations, corresponding with cues given to the subject instructing them to predict the outcome of the top or bottom set of boxes, while four possible outcomes were model, corresponding with feedback to the subject indicating that they should stay or switch for the top and bottom sets.

Four regressors and two parametric modulators were used in GLMs for our model-based analyses. Two regressors modeled left vs. right button presses, as described above. A third regressor, PREDICTION, was modeled as a series of impulse functions at each TR in the interval from the onset of the prediction cue to the delivery of feedback. Finally, the EVALUATION regressor was modeled as an impulse function at the time feedback was delivered. In addition, model-based predictions of neural activity derived from simulations of the PRO model were used to create parametrically modulated PREDICTION and EVALUATION regressors. Participants who committed any errors (N = 8) had two additional regressors included in their analysis – one for the prediction phase on error trials, and one for the outcome phase on error trials. Note that the capitalized words "PREDICTION" and "EVALUATION" refer to different periods within a trial during which subjects will likely be engaging in, respectively, predicting likely outcomes and evaluating observed outcomes. Also, the PREDICTION and EVALUATION regressors are parametrically modulated by the PRO model output, as described below.

Parametric modulators for our model-based analysis were derived from simulations of the PRO model using parameters that were identical to those published previously (Alexander & Brown, 2011), with the exception that each model iteration was interpreted as lasting 100ms, (i.e., each TR corresponded to 20 model iterations). The reason for this change from the original PRO model (Alexander & Brown, 2011) was to allow the model to converge on appropriate predictions given the limited amount of training data (see below). The model was simulated only for the PREDICTION and EVALUATION phases of each trial. Input to the model consisted of two stimuli, corresponding to task cues indicating that the subject would receive feedback related to the top or bottom boxes as described in section 2.2.1. A total of four possible outcomes were modeled: Top/Switch, Top/Stay, Bottom/Switch and Bottom/Stay.

In order to generate parametric modulators for trial-by-trial activity in the behavioral task for a single subject, the PRO model was initially trained on a randomly selected subset of 50 trials (out of 100) that the subject had experienced during scanning. During the training period, weights in the model representing outcome predictions were updated to reflect the model's estimation of the likelihood of observing specific outcomes (see Figure 1 for a conceptual framework of this process). The intended purpose of the training phase was to faithfully replicate the circumstances of our experimental setup in which subjects completed a control run prior to scanning. Following this initial training phase, the model, using the prediction weights obtained during training, was presented with the complete sequence of 100 trials experienced by that subject during scanning in the order in which the subject experienced them. During this sequence, all model learning rules remained in effect. Model activity was recorded on each 100 ms simulation iteration, and was calculated as the rectified value of current, learned predictions

of likely outcomes minus actual outcomes, i.e., negative surprise (Alexander & Brown, 2011), summed over all outcome predictions according to the following equation:

$$Activity_t = \sum_i \left\lfloor Predicted\ Outcome_{i,t} - Actual\ Outcome_{i,t} \right\rfloor^+ \quad (5)$$

where *t* is the current model iteration, and the superscript "+" indicates positive rectification, i.e. that negative values are evaluated as zero. Note that equation (5) is used to compute model activity for both PREDICTION and EVALUATION parametric modulators in the GLM. In the current analyses, we do not model the complement of negative surprise (positive surprise: observed outcomes minus learned predictions) for two reasons. First, a wide range of activity observed in dACC/mPFC has been accounted for using only the notion of negative surprise (Alexander & Brown, 2011); incorporating only the negative component of surprise, therefore, is a more direct test of one of the central claims of the PRO model. Second, positive and negative surprise tend to be directly (though not perfectly) correlated; the absence of a predicted stimulus is often accompanied by the occurrence of an unpredicted stimulus. In the current study, outcomes are binary and are always presented, and so the values obtained from modeling only negative surprise vs. the combination of negative and positive surprise are correlated perfectly. The value of a *Predicted Outcome* is a temporally discounted function reflecting both the learned likelihood of a particular outcome *i* occurring as well as the amount of time until that outcome is expected to occur. On each model iteration, the *Predicted Outcome* is updated to reflect the current time-discounted predicted likelihood of a predicted outcome occurring. The *Actual Outcome* is binary, taking the value of 1 on the model iteration *t* in which a particular outcome is observed, and 0 at all other times. Equation (5) was used to derive parametric modulators for both the PREDICTION and EVALUATION regressors. The parametric modulator for the PREDICTION regressor was calculated for each two second TR as the average model activity of

the 20 iterations starting from the TR onset and ending at the iteration immediately preceding the onset of the next TR. The number of TRs per trial varied due to jitter between the onsets of the PREDICTION phase and EVALUATION phase, ranging from 3 to 7. It may seem counter-intuitive that equation (5) can be used to generate both the PREDICTION and EVALUATION signals, but note that during the PREDICTION interval (prior to the occurrence of an outcome, the value for the *Actual Outcome* is 0 for all *i*, indicating that an outcome has not yet occurred. Model activity during this period therefore reflects only the time-discounted outcome prediction. In this way, equation (5) reflects the PREDICTION signal prior to the outcome, and the EVALUATION signal afterward. The parametric modulator for the EVALUATION phase was calculated as the average activity from equation (5) during the 20 iterations following the delivery of feedback to the model. The procedure described above was conducted twice for each subject's data, once in order to generate PREDICTION modulators, and once in order to generate EVALUATION modulators. The independent simulations were identical with the exception that, for the EVALUATION simulations, the time interval between the beginning of the prediction phase and the delivery of feedback was held constant, while simulations used to generate PREDICTION modulators simulated the jittered interval between the onset of the PREDICTION phase and the EVALUATION phase. The rationale for this is that, due to the procedure used to generate jitter intervals, trials with especially long intervals were severely undersampled due to the relative infrequency of long jitter intervals, resulting in a failure of the model to converge on appropriate predictions regarding outcomes at those times. Additional analysis using EVALUATION modulators generated using simulations incorporating jittered intervals showed effects similar to those reported below, albeit with a substantial loss in power

due to the unreliability of model predictions regarding outcomes following prolonged jitter intervals.

Unless otherwise stated, all results were thresholded at the voxel-level at $p < 0.005$. Cluster extent provided corrections for multiple comparisons ($p < 0.05$ corrected) through AlphaSim (http://afni.nimh.gov/afni/). Based on AlphaSim, whole-brain analyses included a 144 voxel extent criterion.

3.3. Results

Behavioral Results

All participants performed the task at a satisfactory level ($\geq$ 95% correct responses per participant, collapsed across correct switches and correct stays). When errors did occur, there was no significant difference between incorrect switches and incorrect stays ($t(13)=0.849$, $P = 0.404$). However, participants did commit significantly more errors in the *Predict1* condition compared to the *Predict2* condition ($t(13) = 2.09$, $P < 0.05$). Participants were verbally debriefed after the task, and each participant reported that they had understood the task.

Model-Based Results

The PRO model postulates that the mPFC/ACC generates predictions of outcomes, which are then compared against actual outcomes to produce a discrepancy signal that drives future learning (Figure 1). Simulations of the PRO model indicate that the mPFC/ACC should be sensitive to the number of predictions, as well as the degree to which predictions are violated. Here, we explore these effects.

Effect of Prediction

In order to determine whether the mPFC/ACC is sensitive to the number of predictions as anticipated by the PRO model, we regressed neural activity onto the model-based PREDICTION regressors generated by the PRO model, and entered these into the GLM as parametric modulators. A one-sample t-test was then carried out on these (mean-centered) parametric modulators, comparing the effect to a population mean of zero. Confirming model predictions, significant loading on the model PREDICTION signal was found in the left anterior portion of the ACC (MNI -6, 26, 26; k = 484 voxels; peak voxel z-value = 3.49; $P < 0.001$), as well as the caudal cingulate zone (CCZ; MNI -6, -26, 40; k = 12077 voxels; peak voxel z-value = 4.66; $P < 0.001$), as shown in Figure 3. In addition to the ACC, a network of other regions also showed activity consistent with outcome predictions, including the bilateral insula, which has been implicated in outcome prediction (Preuschoff et al., 2008); (Table 1).

Effect of Outcome

Similarly, in order to test which cortical areas are associated with violations of predictions, we tested for areas loading onto the model-based EVALUATION (mean-centered) parametrically modulated regressors generated by the PRO model. Signification loading on the model EVALUATION signal was found in the dorsal ACC (MNI 2, 18, 48; k = 671 voxels; peak voxel z-value = 4.17; $P < 0.001$, cluster corrected). This region was situated in-between areas sensitive to predictions described above, but did not overlap with them (see Figure 3). Other regions loading on the EVALUATION regressors included visual cortex, bilateral insula, bilateral middle frontal gyrus, and left superior frontal sulcus (Table 2). These areas may carry out

separate evaluations of prediction error, as has been shown, for example, in the insula (Preuschoff et al., 2008) and visual cortex (Egner et al., 2010).

Comparison of prediction and surprise effects

It is also possible that the various dorsal ACC regions were simply more active during different phases of the trial as main effects, independent of any parametric modulation by model-based signals. To explore this possibility, an additional contrast was carried out on the un-modulated PREDICTION and EVALUATION regressors to test the overlap of main effects with the parametric modulators. No overlap was found between the un-modulated, main effect of activity during the PREDICTION phase of the trial vs. the parametric modulator for PREDICTION. For the un-modulated, main effect of EVALUATION, overlap was found with the parametric modulator in the dorsal anterior cingulate cortex with the parametric effect of EVALUATION (significant effect of un-modulated EVALUATION REGRESSOR, MNI -2, 6, 48; k = 3854 voxels; peak voxel z-value = 5.47; $P < 0.001$). This overlap is expected because multiple predictions are made in a given trial, so one of them is likely to be violated, which would generally elevate activity in regions that compute prediction error and specifically *negative surprise*. Nevertheless, the parametrically modulated regressors were mean-centered and orthogonal to the corresponding un-modulated regressors, so logically it is possible to see a main effect but not a loading on the parametric modulator in a given region, and *vice versa*. The results overall suggest some main effect of activation in the dorsal ACC during the EVALUATION phase, but this is not confounded with the distinct loading on the PREDICTION and EVALUATION parametrically modulated regressors.

To test furthermore whether specific brain areas were more responsive overall to the un-modulated PREDICTION regressor as opposed to the un-modulated EVALUATION regressor, a paired t-test was carried out to compare beta estimates for EVALUATION relative to PREDICTION. This comparison revealed significantly greater activity during the EVALUATION phase in the dorsal ACC (MNI 4, 12, 46; k = 4133 voxels; peak voxel z-value = 5.29; $P < 0.001$), consistent with previous results showing generally strong activity in this region during outcome relative to prediction (Brown, 2009; Jahn et al., 2011). The opposite contrast of PREDICTION minus EVALUATION showed activation mainly in white matter regions.

Dissociation of prediction and surprise effects

In order to test whether each of the ACC regions were preferentially activated to only one of the contrasts and not the other, unbiased ROIs were created in order to test for a significant ROI x Contrast interaction (Nieuwenhuis et al., 2011). Two spherical ROIs (5mm each) were placed in distinct ACC subregions outlined by Nee et al (2011) to demarcate the structural and functional heterogeneity of the ACC without being near enough to have parameter estimates from each ROI unduly affected by smoothing (Figure 3). The first ROI was placed in the posterior rostral cingulate zone (RCZp; MNI center 0, 10, 46). The second spherical ROI was placed in the rostral cingulate gyrus (CG; MNI center 0, 38, 10). An additional 5mm spherical ROI was placed in the caudal cingulate zone (CCZ; MNI center, 0, -10, 39), in order to extend the coverage of our analysis to the posterior regions of the cingulate. The location of this ROI was taken from peak voxel coordinates for a contrast of strong vs. medium anticipation (Drabant et al., 2011). These three regions were selected to serve as unbiased ROIs corresponding to known functional and anatomical subdivisions within the ACC (Fan et al., 2008; Paus, 2001)

A significant ROI x Condition interaction was found ($F(2, 26) = 3.75$, $P < 0.05$), driven by greater effects for EVALUATION than PREDICTION in region RCZp, and the reverse pattern (greater effects for PREDICTION than EVALUATION) in both CG and CCZ (Figure 3). Within each ROI, paired-t-tests were conducted to test for significant differences between the effects of PREDICTION and EVALUATION. Bonferroni correction for multiple comparisons was used when comparing mean differences, resulting in a corrected critical t-value of 2.75. Within RCZp there was a significant effect of EVALUATION ($t(13) = 3.54$, $P < 0.01$) and a non-significant result of PREDICTION ($t(13) = 0.34$, $P > 0.05$), with a paired t-test between the conditions showing no significant difference ($t(13) = 2.32$, $P < 0.05$). The opposite pattern was found within CCZ with a significant effect of PREDICTION greater than EVALUATION ($t(13) = 3.13$, $P < 0.01$), driven by a significant effect of PREDICTION ($t(13) = 4.30$, $P < 0.01$) and a non-significant result of EVALUATION ($t(13) = 0.45$, $P > 0.05$). Within CG, there was a trend towards a significant effect of PREDICTION ($t(13) = 2.37$, $P < 0.1$) but no effect of EVALUATION ($t(13) = 0.76$, $P > 0.05$), although a paired t-test revealed no significant difference between the parameter estimates ($t(13) = 1.31$, $P > 0.05$). Overall, these results lend support to the proposal that distinct sub-regions of the ACC are involved in prediction and outcome calculations, which is consistent with the PRO model.

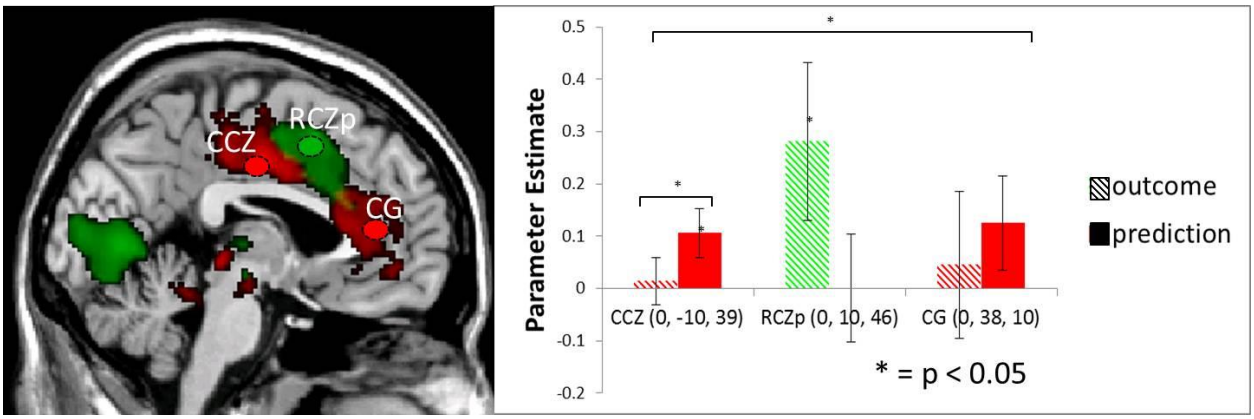

Figure 3: Regions of the CCZ/ACC showing increased activation in response to prediction- and outcome-related effects as predicted by the PRO model, presented at a cluster corrected threshold of $p < 0.05$. The RCZp and CG ROIs were taken from Nee et al (2011), while the CCZ ROI was created from coordinates reported in Drabant et al (2011). Red: Prediction-related effects; Green: Outcome-related effects.

Table 1

Brain Region (TD)	X	Y (MNI)	Z	Z-score	Cluster Corrected p-value	Cluster Size
Effect of model PREDICTION regressor						
Left Supramarginal Gyrus	-56	-24	42	5.75	<0.001	12077
Caudal Cingulate Zone*	-6	-26	40	4.66	<0.001	
Left Thalamus	-14	-32	0	4.66	<0.001	399
Right Anterior Insula	34	32	-12	4.58	<0.001	929
Left Insula	-48	22	2	3.83	<0.001	950
Left Rostral ACC	-6	26	26	3.49	<0.001	484

Left Inferior Parietal Lobe	-46	-60	10	3.42	<0.05	173

Table 2

Brain Region (TD)	**X**	**Y (MNI)**	**Z**	**Z-score**	**Cluster Corrected p-value**	**Cluster Size**
Effect of model EVALUATION regressor						
Visual Cortex	6	-64	0	4.85	<0.001	2509
Left Superior Frontal Sulcus	-26	0	52	4.36	<0.05	202
Dorsal ACC	2	18	48	4.17	<0.001	671
Right Hippocampus	18	-26	-8	4.12	<0.001	531
Left Anterior Insula	-32	22	-6	4.08	<0.01	271
Right Frontal Middle Gyrus	46	6	30	3.99	<0.01	311
Left Frontal Middle Gyrus	-42	-2	24	3.96	<0.05	189
Right Thalamus	10	-20	14	3.92	<0.01	288
Right Anterior Insula	36	18	2	3.59	<0.05	168
Left Parietal Inferior Lobe	-48	-38	38	3.24	<0.05	149

Control Analysis

A control run was presented to thirteen subjects, in which the same prediction phase images were presented to the subjects, but no responses were made and no predictions formed. Within the independent ROIs of CCZ and CG, parameter estimates were extracted for the contrast *Predict2-Predict1* of the control run, and compared to parameter estimates for *Predict2-Predict1* of the experimental runs. Within CCZ, the *Predict2-Predict1* contrast for the experimental runs was significantly greater than zero ($t(12) = 3.49$, $P < 0.01$), while the same contrast for the control run was not significantly greater than zero ($t(12) = -0.29$, $P = 0.77$). For region CG, on the other hand, the *Predict2-Predict1*contrast was significantly greater than zero ($t(12) = 3.27$, $P < 0.01$), while prediction-related activity for the control run was significantly less than zero ($t(12) = -3.59$, $P < 0.01$; Figure 4). These results suggest that prediction-related activity was not solely driven by oculomotor or attention-related processes.

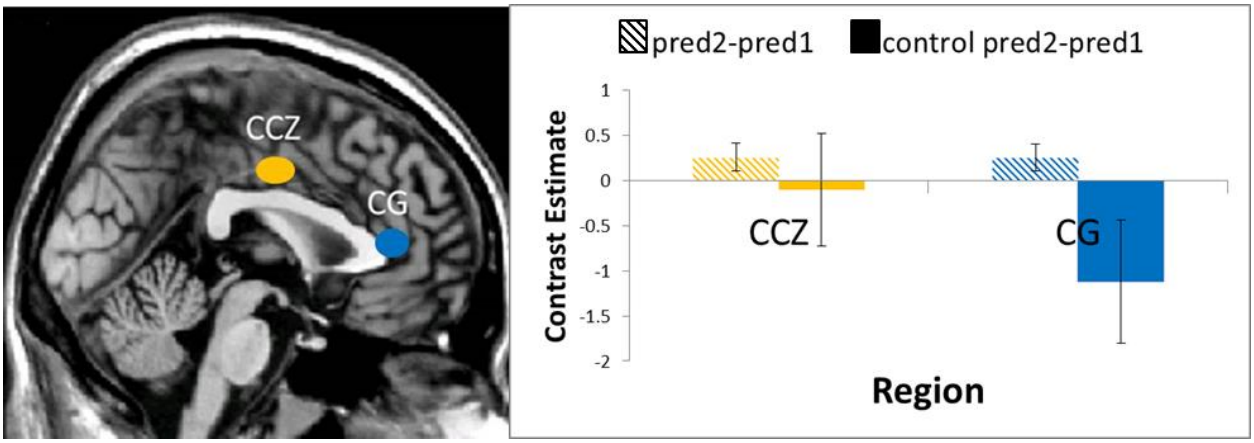

Figure 4: Regions of the posterior cingulate (CCZ) and anterior cingulate (CG) showing increased activation for predicting 2 outcomes (*Predict2*) as opposed to predicting 1 outcome (*Predict1*). Within these independent ROIs, parameter estimates were extracted for the control condition. A paired t-test in both regions revealed that the parameter estimates for the *Predict2-Predict1* contrast were significantly greater than for the control contrast. The control condition did not pass any corrected thresholds in a whole-brain analysis.

Chapter 4

Prediction error across domains in the mPFC

The previous two neuroimaging studies have provided a substantial foundation for both comparing the PRO model against the conflict monitoring account of mPFC function (Jahn et al, 2011), as well as directly applying the PRO model to generate regressors to distinguish between prediction- and outcome-related processes within the mPFC (Jahn et al, 2014) without altering the published PRO model parameters of Alexander & Brown, 2011. However, one critical aspect of the PRO model that remains to be tested is whether the mPFC computes a modality-general prediction error signal, and whether this can be separated anatomically from the processing of other affectively salient information, such as pain, which nevertheless can be segregated from prediction error processes.

As discussed above, the ACC has been shown to be highly responsive to both the receipt of painful and aversive stimuli (Derbyshire et al., 1998), as well as when an individual encounters a situation requiring a high level of cognitive control (Botvinick et al., 2004). However, to our knowledge there has not yet been an experiment which that has directly compared both the prediction and outcome phases of these different conditions. A combination of both conditions in a single paradigm would allow for the direct comparison of prediction-related activity of aversive stimulation to prediction-related activity for cognitive control while matching for levels of prediction and deviation from that prediction, and would provide further insight into how cortical areas such as the ACC and neighboring regions of medial PFC are involved in the prediction of qualitatively different levels of aversive stimuli. The results would also allow the adjudication between competing models of ACC activity – specifically the reinforcement learning model (Holroyd & Coles, 2002) and the PRO model (Alexander &

Brown, 2011) – and provide a stronger theoretical framework for interpreting the results of the nicotine study discussed above. According to the reinforcement learning model, the ACC should be more sensitive to negatively valenced outcomes, regardless of how unexpected they are. The PRO model, on the other hand, predicts that both negatively and positively valenced outcomes matched on unexpectedness – i.e., prediction error – should elicit similar patterns of activity in the ACC. The current study will match both types of outcomes on their prediction error.

Furthermore, in light of the findings by Aarts and colleagues (2011) that the ACC responds to the probability of receiving an upcoming incongruent Stroop task, and based on other experiments showing that the ACC is responsive to the unexpected absence of pain (Chandrasekhar et al., 2008), the proposed study will combine both of these conditions in a factorial design. This is motivated by the PRO model hypothesis that the ACC acts as an action-outcome predictor, and will attempt to dissociate whether distinct neural populations within the ACC are responsible for the generation of the same prediction signals for both pain and for cognitive control, or whether these are located within distinct regions of the ACC.As it stands, either homogenous or heterogeneous localization of prediction error within the mPFC would be consistent with the PRO model, as this experiment represents the first empirical test of whether distinct prediction error regions exist within the mPFC. Examining the location of prediction error for both cognitive and affective modalities (represented by Stroop and shock stimuli, respectively) would further refine the PRO model and understanding the functional architecture of the prefrontal cortex.

4.1. Participants

Data from 29 right-handed participants (10 female) were collected (mean age = 24.0, SD = 2.80). Participants reported no history of psychiatric or neurological disorder, and reported no current

use of psychoactive medications. Participants were compensated $30/hour for their time. Participants were trained on the task on a computer outside of the scanner for two practice blocks on a separate day from the scanning session. When they returned for the scanning session, participants were run on one practice block of the task outside of the scanner before undergoing the scanning session. Data from two subjects were discarded due to insufficient accuracy (more than three standard deviations below the mean accuracy of all subjects), and data from an additional subject was discarded due to a self-reported failure to follow the instructions, leaving a total of 26 useable participants. Within this sample, useable galvanic skin response (GSR) signal was acquired for 23 subjects.

4.2. Experimental Paradigm

Electrical Stimulation Apparatus

To deliver electrical stimulation, a transcutaneous aversive finger stimulator was used (Model E13-22, Coulbourn Instruments). The range of electrical stimulation delivered by the device ranged from 0.2mA to 4.0mA, with nine discrete steps. MRI-compatible electrodes were placed on the pinky and ring fingers of the left hand. Before undergoing scanning, participants were administered the lowest possible level of shock from the finger stimulator. The current was raised incrementally, and participants were instructed to tell the experimenter when the current had reached a level of stimulation that was "Aversive, but not painful". This setting was used as their high level of electrical stimulation. Starting again from the lowest level of stimulation, participants were instructed to tell the experimenter when the level of stimulation was noticeable, but not aversive. This setting was used as their low level of electrical stimulation.

During the scanning session at the end of each experimental block, participants were asked whether either level of stimulation was too high or too low. The levels of stimulation were then adjusted until the participant reported that both the high level and low level of electrical stimulation met the original criteria. The lowest setting chosen by participants was 0.5mA, while the highest setting selected as aversive without being painful was 2.7mA.

Galvanic Skin Response (GSR)

GSR data was collected using the MP-150 system (BIOPAC Systems Inc., CA, USA) at a sampling rate of 250 Hz by using MRI compatible electrodes placed on the thenar and hypothenar of the left hand. GSR data were lowpass filtered allowing frequencies below 15Hz and detrended by subtracting the mean for the entire run from each sampled datapoint before extracting the peak GSR value within a window of 1-6s after the administration of electrical stimulation.

fMRI Paradigm and Model Regressors

The task was designed to compare cognitive and affective surprise. To achieve this, a 2x3 factorial design was used. The factor of modality consisted of two levels: Stroop and pain. The cue factor consisted of three levels: a cue signaling a 75% chance of obtaining an aversive outcome (incongruent stimulus or more painful stimulus), a cue signaling a 75% chance of obtaining a non-aversive outcome (congruent stimulus and less painful stimulus), and a cue signaling a 50% chance of obtaining either an aversive or non-aversive outcome.

The combination of these factors led to six prediction-related regressors: Three related to pain trials (CueLo, CueHi, CueEitherPain), and the other three related to conflict trials (CueCon,

CueInc, and CueEitherStroop). The outcome phase for the pain condition consisted of either a high-level, aversive shock, or a low-level, non-aversive shock. The outcome phase for the Stroop condition consisted of either an incongruent spatial Stroop stimulus (e.g., the word "Left" printed inside of an arrow pointing to the right), or a congruent spatial Stroop stimulus (e.g., the word "Right" printed inside of an arrow pointing toward the right). These combinations lead to an additional six regressors for outcome in the pain condition – loLo (in which "lo" refers to the low probability of pain condition, and "Lo" refers to the actual outcome of low pain), loHi, hiLo, hiHi, eitherHi, and eitherLo – and an additional six regressors for outcome in the Stroop condition – conCon, conInc, incCon, incInc, eitherCon, and eitherInc.

This combination of prediction cues and actual outcomes led to either better-than-expected or worse-than expected outcomes. For the pain condition, for example, a worse-than-expected outcome would consist of receiving a prediction cue associated with a 75% chance of obtaining a non-aversive shock, and then during the outcome phase receiving an aversive shock. Similarly, a worse-than-expected outcome for the Stroop condition would consist of receiving a prediction cue representing a 75% chance of obtaining a congruent spatial Stroop, but then receiving an incongruent spatial Stroop (see Figure 1).

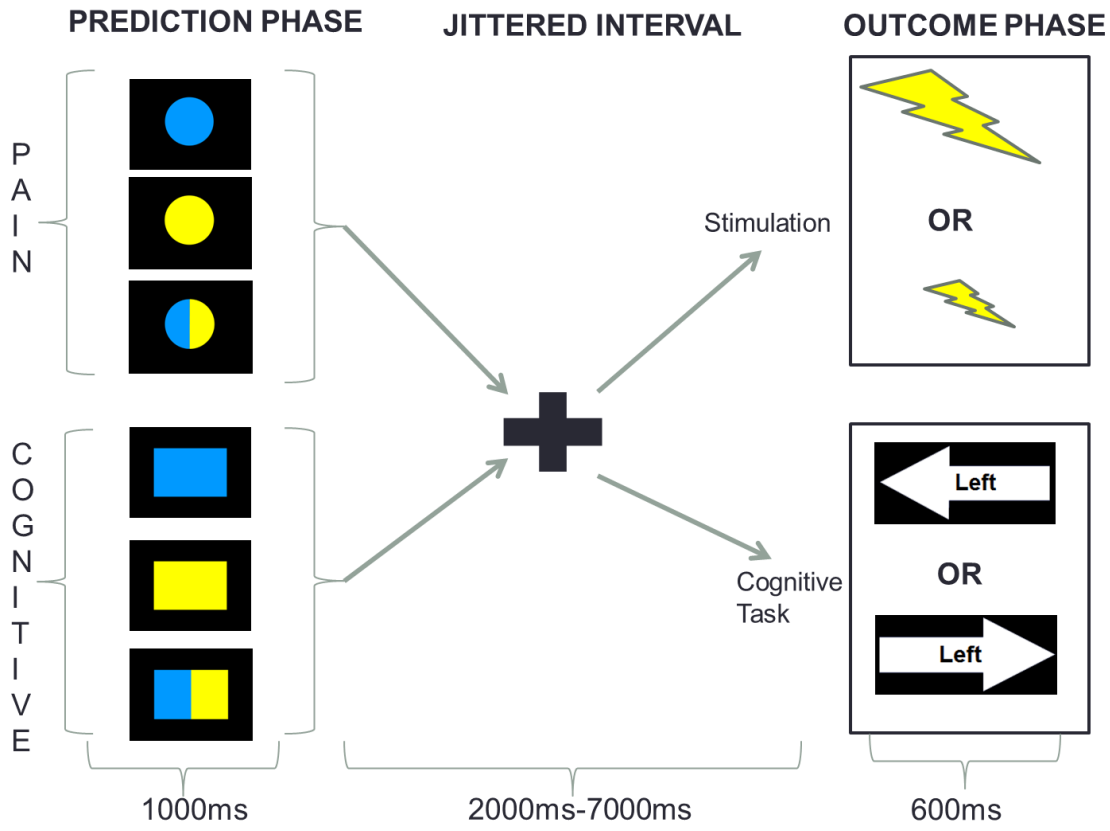

Figure 1: Task design. During the prediction phase subjects are presented with a cue signaling whether the outcome will be pain or a spatial Stroop, as well as the probability of whether it will be a better-than-expected or worse-than-expected outcome. Cues that are half blue and half yellow always signify that there is an equal chance of receiving either a better-than-expected or worse-than-expected outcome; the probabilities signified by either blue or yellow cues was counterbalanced across subjects. Although this figure depicts circles as signaling the electrical stimulation condition and rectangles signaling the spatial Stroop condition, the meaning of the shapes was also counterbalanced across subjects. The prediction phase is followed by a jittered interval, and then the outcome of either an electric shock or a spatial Stroop task

The contrast of cognitive prediction errors contrasted conditions where predictions were violated against conditions where predictions were met [conInc + incCon] – [conCon + incInc]. Similarly, the contrast for affective prediction errors contrasted prediction violations about pain against outcomes where the prediction was met [loHi + hiLo] – [hiHi + loLo].

The prediction phase for each trial lasted for 1s, during which participants were presented with a cue to predict the outcome. This prediction phase was then followed by a jittered interval of 2-7swhere the participants only saw a fixation cross. The outcome phase lasted for 600ms, during which either a Stroop trial was presented or the electrical stimulation was administered. This was followed by another jittered interval of 4-8s, to allow for independent estimation of the BOLD signal between the two prediction phase and outcome phase.

In total there were five runs of scanning per subject. Each run contained 36 Stroop trials and 36 pain trials, for a total of 180 trials for each modality per subject. There was an even number of trials using aversive predicting, non-aversive predicting, and uninformative cues. For aversive and non-aversive predicting cues, 25% of the outcomes of the trials were better-than-expected or worse-than-expected, while 75% of the outcomes were as-expected from the predictive cue. For uninformative cues, 50% of the outcomes were aversive, while 50% of the outcomes were non-aversive. In total there were 15 better-than-expected and 15-worse-than-expected outcomes for each modality.

All of these regressors, along with a separate regressor modeling the reaction time for each trial, were estimated in a GLM referred to here as GLMNOGSR. For another one of the GLMs, referred to hereafter as GLMGSR, GSR was included as the sole regressor, along with motion regressors if the participant's motion exceeded 3mm in any direction or rotation for the session. As GSR was sampled every TR there was the possibility that a model including other regressors in addition to GSR would be biased in assigning variance to GSR only. For this GLM each participant's GSR data was convolved with a canonical hemodynamic response function and then subsampled every 3 seconds. This resulting GSR timecourse was then inserted as a regressor into the model. We were able to use this GLM for the 23 subjects who had useable

GSR. Unless otherwise stated, the analyses described here were taken from the GLMNOGSR model.

Leave-One-Out Analysis

To create data-driven, independent ROIs to examine each effect, we used a leave-one-out cross-validation method (Esterman et al, 2010). In this approach second-level analyses are run for each contrast, consecutively leaving out each subject and extracting contrast estimates from the resulting ROI.

4.3 fMRI analysis

Image acquisition and preprocessing

The experiment was conducted with a 3 Tesla Siemens Trio scanner using a 32-channel head coil. Foam padding was inserted around the sides of the head to increase participant comfort and reduce head motion. Imaging data was acquired at a 30° angle from the anterior commissure-posterior commissure line (Deichmann, Gottfried, Hutton, & Turner, 2003). Functional T2* weighted images were acquired using a gradient echo planar imaging sequence [50 x 2.7mm interleaved slices; TE = 25ms; TR = 3000ms; 96x96 voxel matrix; 220x220mm field of view]. For the experimental condition, five runs of data were collected with 208 functional scans each. High resolution T1-weighted images for anatomical data [256x256 voxel matrix] were collected at the end of each session.

Functional data were spike-corrected using AFNI's despiking algorithm (http://afni.nimh.nih.gov/afni). SPM5 (Wellcome Department of Imaging Neuroscience, London, UK; www.fil.ion.ucl.ac.uk/spm) was used for subsequent preprocessing and data analysis. The

functional data for each run for each participant was slice-time corrected and realigned to each run's mean functional image using a 6 degree-of-freedom rigid body spatial transformation. The resulting images were then coregistered to the participant's structural image. The structural image was normalized to standard Montreal Neurological Institute (MNI) space and the transformations were applied to the functional images. The functional images were then spatially smoothed using an 8mm Gaussian kernel. Regressors were modeled with an impulse delta function convolved with the hemodynamic response function at the time of onset. Each type of cue for both Stroop trials and pain trials were modeled at the onset of the cue, as well as the outcome for both Stroop trials and pain trials.

Unless otherwise stated, all results were thresholded at the voxel-level at $p < 0.01$. Cluster extent provided corrections for multiple comparisons ($p < 0.05$ corrected) through AlphaSim. Based on AlphaSim, whole-brain analyses included a 360 voxel extent criterion.

4.4 Results

Behavioral Results

Subjects performed the task at a satisfactory level, and behavioral performance could only be measured in the Stroop task (mean percentage correct = 87.1%; SD = 10.2%. Range = 65%-98%). Participants were verbally debriefed after the task, and all participants indicated that they had understood the task and followed the instructions.

The typical timecourse of the GSR response is a delayed rampup beginning about 1-2 seconds after the onset of the shock, and peaking approximately 2-4 seconds after the onset of the shock, gradually falling back to baseline (see Figure 2 for a representative example for a single subject.) To confirm that high intensity shocks were indeed more aversive than low

intensity shocks, we examined the GSR data. Consistent with this idea, a paired t-test revealed a significantly higher GSR for the high-shock events than the lower shock events (M for high-shock events = 0.87mS; M for low-shock events = 0.01mS; $t(22)=2.79$, $P < 0.05$.) Additional GSR analyses revealed that the signal collapsed across each of the Stroop conditions was significantly greater than zero ($M = 0.019$mS; $t(22) = 8.01$, $P < 0.05$); however, none of the Stroop conditions were significantly different from each other (see Figure 2A).

A 2 (task) x 3 (prediction) x 2 (outcome) ANOVA was carried out to examine differences in GSR. There was a trend toward a main effect of task ($F(1, 22)=3.75$, $P = 0.07$) as well as a main effect of outcome ($F(3, 66)=7.92$, $P < 0.01$). There was a significant task x outcome interaction ($F(1,22) = 5.13$, $P < 0.05$), and a significant task x outcome x prediction interaction ($F(2, 44) = 7.71$, $P < 0.01$).

A 2x3 ANOVA was carried out to examine differences and interactions of reaction times across the Stroop conditions. A comparison of reaction times across conditions revealed a significant main effect of incongruent Stroop trials as compared to congruent Stroop trials, collapsed across levels of expectation ($F(1,25) = 14.21$, $P < 0.001$)) but no effect of prediction ($F(1,25)=0.44$, $P = 0.65$). In addition, there was a significant interaction between prediction and outcome conditions ($F(2,50 = 5.25$, $P < 0.01$)). The interaction was driven by faster reaction times for outcomes that were consistent with predictions, and slower reaction times for outcomes that were inconsistent with predictions (Figure 2B), replicating previous findings (Aarts et al., 2011).

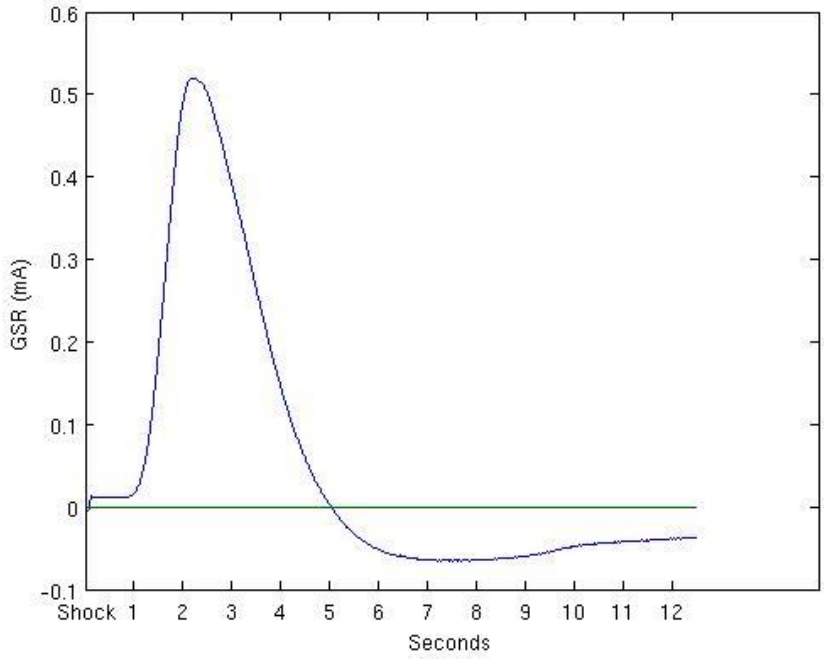

Figure 2: Typical GSR timecourse for single subject. Onset of shock is labeled at the beginning of the x-axis, showing a rampup of approximately 2 seconds and a peak at around 3 seconds after the onset of shock.

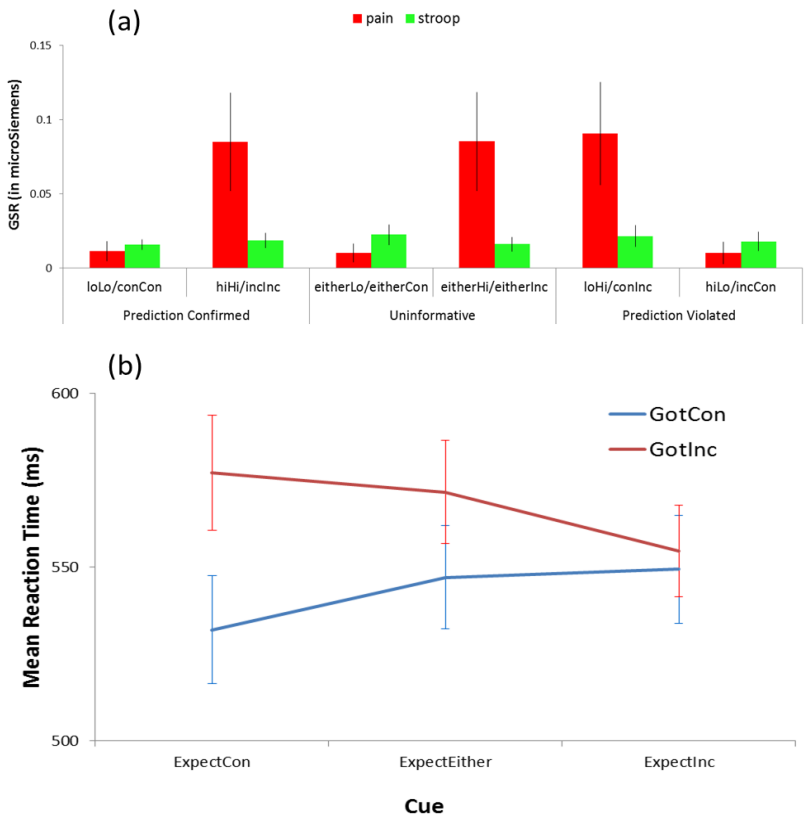

Figure 3: (A) Amount of GSR in microSiemens (mS) for cue and outcome conditions in both the pain and Stroop tasks. A main effect of high pain was found ($t(22) = 2.79$, $P < 0.05$); none of the other contrasts were significant. Error bars represent one SEM. (B) RTs across Stroop conditions, in seconds. Error bar represents two standard errors of the mean. The interaction between the conditions was found to be significant, $P<0.05$.

Imaging Results

For the main effect of pain, a significant cluster of activation was found within the rostral ACC (MNI -2 30 14; k = 1753 voxels; peak z-value = 5.17; $P < 0.01$, cluster corrected), consistent with this region's role in processing pain and negative affect (see Figure 3). Next we looked for a main effect of incongruency (i.e. conflict effects) by contrasting incongruent Stroop trials with congruent Stroop trials. However, no effects were found for this contrast.

Looking at only incongruent trials after congruent Stroop cues, compared to congruent trials after congruent cues, we found a significant cluster in the dACC (MNI 0 20 44; k = 536 voxels; peak z-value = 3.62; $P < 0.001$, cluster corrected), similar to the same contrast reported in Aarts et al (2011).

Multi-modal surprise in preSMA: Cognitive vs. Affective prediction errors

First, we examined the effect of pain prediction error collapsing across valence. This contrast revealed activation in the preSMA/dACC area (MNI -4, 20, 48; k = 1076 voxels; peak voxel z-value = 3.71; $P < 0.01$, cluster corrected). A similar contrast was carried out with surprising effects of Stroop, collapsing across valence, which revealed a significant cluster in the dorsal ACC (MNI 6, 18, 42; k = 878 voxels; peak voxel z-value = 4.15, $P < 0.001$, cluster corrected), as shown in Figure 3A.

Next, we conducted a slice-by-slice analysis on the Stroop surprise and pain surprise clusters by taking the main effect of prediction error across both Stroop and pain conditions. Contrast estimates were then averaged across each slice along the anterior-posterior axis, with each slice defined by its corresponding Y-axis coordinate. A test of these contrast estimates revealed a significant interaction of condition by slice ($F(1,25) = 5.07$, $P < 0.05$), as shown in Figure 3B. This analysis suggests that despite substantial overlap of pain and Stroop surprise effects, they are not completely overlapping, which means that Pain and Stroop surprise are represented distinctly in the medial wall.

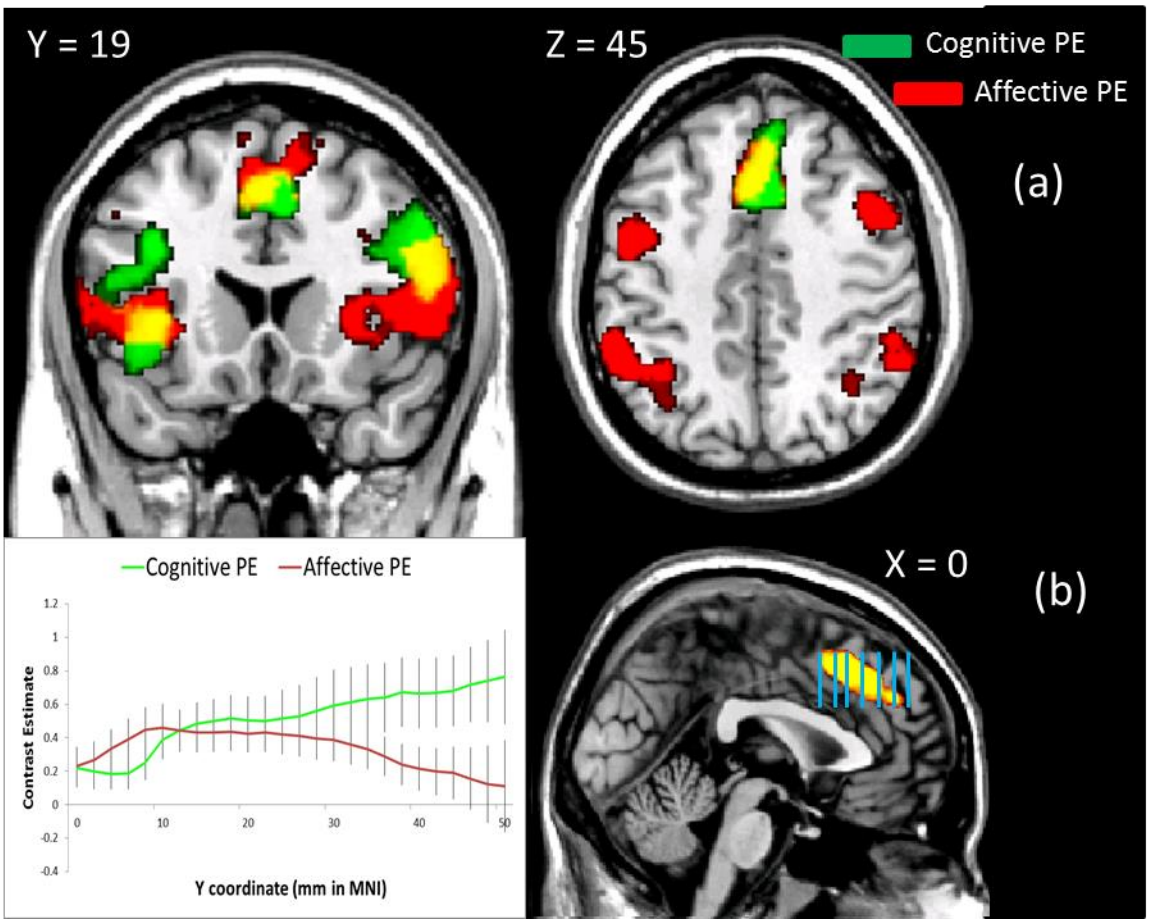

Figure 4: (A) Affective PE vs. Stroop Cogntiive PE**.** Results contrasts for Cognitive PE (green; MNI 8, 24, 38) and Affective PE (red; MNI -4 24 46), depicted at a threshold of $p < 0.05$, cluster corrected. Yellow: Overlap between contrasts of Cognitive PE and Affective PE. (B) Slice analysis of contrast estimates. An OR mask was generated by combining voxels for both the contrasts of Affective PE and Cognitive PE, thresholded at $p < 0.01$ and passing cluster correction at $p < 0.05$. Contrast estimates were averaged across each 2mm slice from posterior to anterior for each subject. A position x condition interaction was found, $F(1, 25) = 5.07$, $P < 0.05$.

Table 1

Brain Region (TD)	X	Y (MNI)	Z	Z-score	Cluster Corrected p-value	Cluster Size
PainSurprise						
Left Anterior Insula	-46	12	4	4.44	<0.01	1589
Left Inferior Parietal Lobule	-64	-42	24	4.33	<0.01	1348
Right Inferior Parietal Lobule	52	-38	38	4.10	<0.01	1082
Right Anterior Insula	46	22	8	3.83	<0.01	1918
preSMA/dorsal ACC	-4	20	48	4.34	<0.01	1076

Table 2

Brain Region (TD)	X	Y (MNI)	Z	Z-score	Cluster Corrected p-value	Cluster Size
StroopSurprise						
Left Inferior Frontal Gyrus	-38	20	26	3.97	<0.01	1349
Right Middle Frontal Gyrus	54	18	32	3.61	<0.01	1382
Dorsal ACC	6	18	42	4.15	<0.01	878

Overlapping Regions Represent Both nPE and pPE Pain

To examine whether both positive and negative PE's have a common neural source, we separately examined negative and positive PE's. For nPE pain, we observed a cluster of activation in the bilateral dACC/pre-SMA region (MNI 4, 38, 40; k = 668 voxels; peak voxel z-

value = 3.39; p < 0.01, cluster corrected), and a comparison contrast of pPE pain revealed activation in the bilateral SMA (MNI 12, 6, 56; k = 610 voxels; peak voxel z-value = 3.85; p < 0.01, cluster corrected). These activations partially overlapped, as shown in figure 4A.

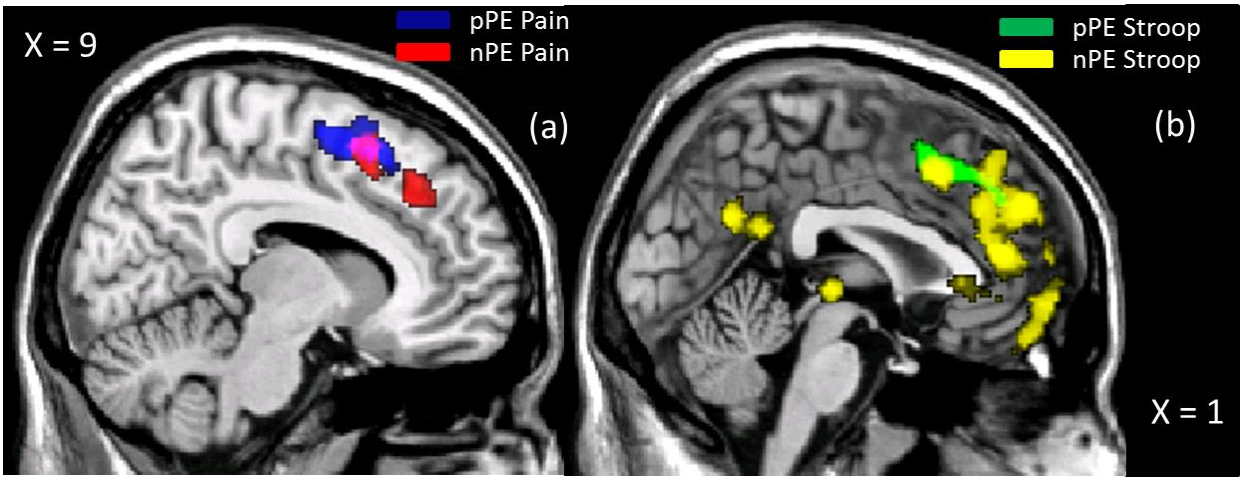

Figure 4: Better vs. worse than expected outcomes second-level results displayed on template MNI152 brain. (A) Better and worse pain surprise representations overlap in preSMA. Results of pPE pain (blue) and nPE pain (red), depicted at a threshold of p < 0.01, cluster corrected. (B) Better and worse Cognitive PE representations overlap only modestly. Results of pPE Stroop (green) and nPE Stroop (yellow), depicted at a threshold of p < 0.05, cluster corrected.

However, it is possible that within these surprising outcomes, negatively valenced outcomes can be weighted more than positively valenced outcomes, even when matched for their level of prediction error, as would be predicted by the reinforcement learning model (Holroyd & Coles, 2002). To test, this, we took a 5mm dACC ROI from Holroyd et al (2004), MNI 1, 18, 44, there was no difference between nPE and pPE Pain (nPE-pPE) ($t(25) = -1.77$, $p = 0.09$). There was an effect of pPE Pain ($t(25) = 2.43$, $p = 0.02$) but no effect of nPE Pain ($t(25) = 0.50$, $p = 0.62$).

Neural Representations of nPE and pPE Stroop Outcomes Partially Overlap

We then split apart the Stroop surprise contrast into both positive prediction error (pPE; incCon – conCon) and negative prediction error (nPE; conInc – incInc) Stroop outcomes. The contrast for pPE Stroop outcomes revealed activation in the dACC (MNI 6, 18, 42; k = 589 voxels; peak voxel z-value = 3.59, $p < 0.01$, cluster corrected). The contrast for nPE Stroop outcomes, on the other hand, did not show any significant clusters at our threshold. At a more liberal voxelwise threshold of $P < 0.05$, a significant cluster was found which encompassed both the dACC and more rostral prefrontal cortex (MNI 4, 52, 32; k = 2096 voxels; peak voxel z-value = 3.07, $p < 0.05$, cluster corrected). Notably, the pPE Stroop contrast revealed activation in the same area of SMA as the affective PE contrast, as depicted in Figure 4B. These results suggest that dACC computes prediction error, and that multiple modalities of surprise are represented in an overlapping manner. In contrast to the mostly overlapping representations of nPE and pPE pain, nPE and pPE Stroop outcomes showed activation patterns in mostly distinct regions of the mPFC. pPE Stroop, as reported above, was found in the SMA overlapping with surprising pain, while nPE Stroop outcomes were located in the rostral cingulate cortex (rACC). Using a 5mm dACC ROI from Holroyd et al (2004), MNI 1, 18, 44, there was no difference between nPE and pPE Stroop (nPE-pPE) ($t(25) = -1.491$, $p = 0.15$). There was no effect of nPE Stroop ($t(25) = -0.47$, $p = 0.64$) or pPE Stroop ($t(25) = 1.66$, $p = 0.11$).

It is possible that the observed results for the Stroop contrasts could be driven by differences in cognitive engagement. For example, the more anterior region overlaps with the so-called "default" network (Fox et al., 2005) and may reflect cognitive disengagement in response to an unexpectedly easy trial. By contrast, the more posterior region overlaps with the so-called "salience" network (Menon et al, 2010) and may reflect the need for cognitive engagement in

response to an unexpectedly difficult trial. If this is true, one might expect the more anterior region to be generally more responsive to congruent than incongruent trials, with the opposite being true of the more posterior region.

To test this, parameter estimates for incongruent and congruent trials were extracted from the nPE cluster in the rACC and the pPE cluster in the SMA, and submitted to a 2 (Congruency) x 2 (ROI) ANOVA. There was a main effect of ROI ($F(1, 25) = 28.43$, $p < 0.001$), with the SMA showing higher parameter estimates for both congruent and incongruent trials. However, there was no main effect of congruency collapsing across regions ($F(1, 25) = 1.337$, $p = 0.259$), and there was no significant interaction term ($F(1, 25) = 0.18$, $p = 0.675$).

Cue Prediction Analysis

Previous studies on prediction-related effects of cues have shown mixed results. In particular, Aarts et al (2008) reported greater activation within the ACC for informative cues contrasted against uninformative cues, but a follow-up study with probabilistic cues – much like the ones used in the current study – showed no whole-brain effects, as well as no significant differences in the unbiased ROI used from the previous study (Aarts et al, 2011).

To test this we carried out a whole-brain analysis of informative cues contrasted against uninformative cues in both the pain and Stroop conditions. For the pain condition, activation was observed only in bilateral visual cortex (MNI 14, -44, 2; k = 1696 voxels; peak voxel z-value = 4.44; $P < 0.001$; and MNI -14, -58, -10; k = 582 voxels; peak voxel z-value = 4.40; $P < 0.001$). No significant clusters were observed within the gray matter for the Stroop condition.

To examine whether there may be prediction-related activation that was missed due to thresholding, we examined cue-related activations in a number of ROIs defined by previous

studies. These ROIs included the ACC ROI from Aarts et al (2008) resulting from the contrast of predictive cues – unpredictive cues, and within the prediction-sensitive ROIs from Jahn et al (2014) within the rostral ACC and caudal cingulate zone (CCZ). For the Aarts et al (2008) ROI, a 5mm sphere was placed at MNI coordinates -9, 13, 41; while for the Jahn et al (2014) ROIs, 5mm spheres were placed at MNI coordinates 0, -10, 39 for the CCZ ROI and MNI coordinates 0, 38, 10 for the rostral ACC ROI. Contrast estimates were extracted from each of these ROIs for the painPrediction and StroopPrediction contrasts described above. No region demonstrated an effect of painPrediction (all $p > 0.5$) nor StroopPrediction (all $p > 0.1$). Thus, there was no evidence for prediction-related activation during the cue phase in these data.

Separate Regions of mPFC Process PE/Conflict and Pain/GSR Effects

ROIs created using the leave-one-out procedure were used to test for potential regional dissociations between effects of PE, Conflict, Pain, and GSR. Using a 2 (Prediction) x 2 (Outcome) x 4 (ROI) ANOVA, interaction effects were found for conflict ($F(3, 75)=4.44$, $p<0.01$), pain ($F(3,75)=8.4$, $p<0.01$), cognitive PE ($F(3,75)=11.8$, $p <0.01$), and affective PE ($F(3,75) = 9.13$, $p <0.01$), as well as main effects of ROI for conflict ($F(3,75) = 8.18$, $p<0.01$), pain ($F(3,75) = 6.42$, $p<0.01$), cognitive PE ($F(3,75) = 6.33$, $p<0.01$), and affective PE ($F(3, 75) = 9.85$, $p < 0.01$). In addition, using these same ROIs to extract GSR from the GLMGSR contrast maps, there was an effect of ROI on GSR ($F(3, 66) = 3.11$, $p<0.05$). These results were driven by stronger PE and conflict effects in the more dorsal mPFC and weaker Pain effects, while this pattern was reversed in the ventral area of the ACC. From these results, it appears that a more dorsal region of the medial PFC processes conflict and surprise, while a more ventral portion of the ACC processes pain and GSR (Figure 5).

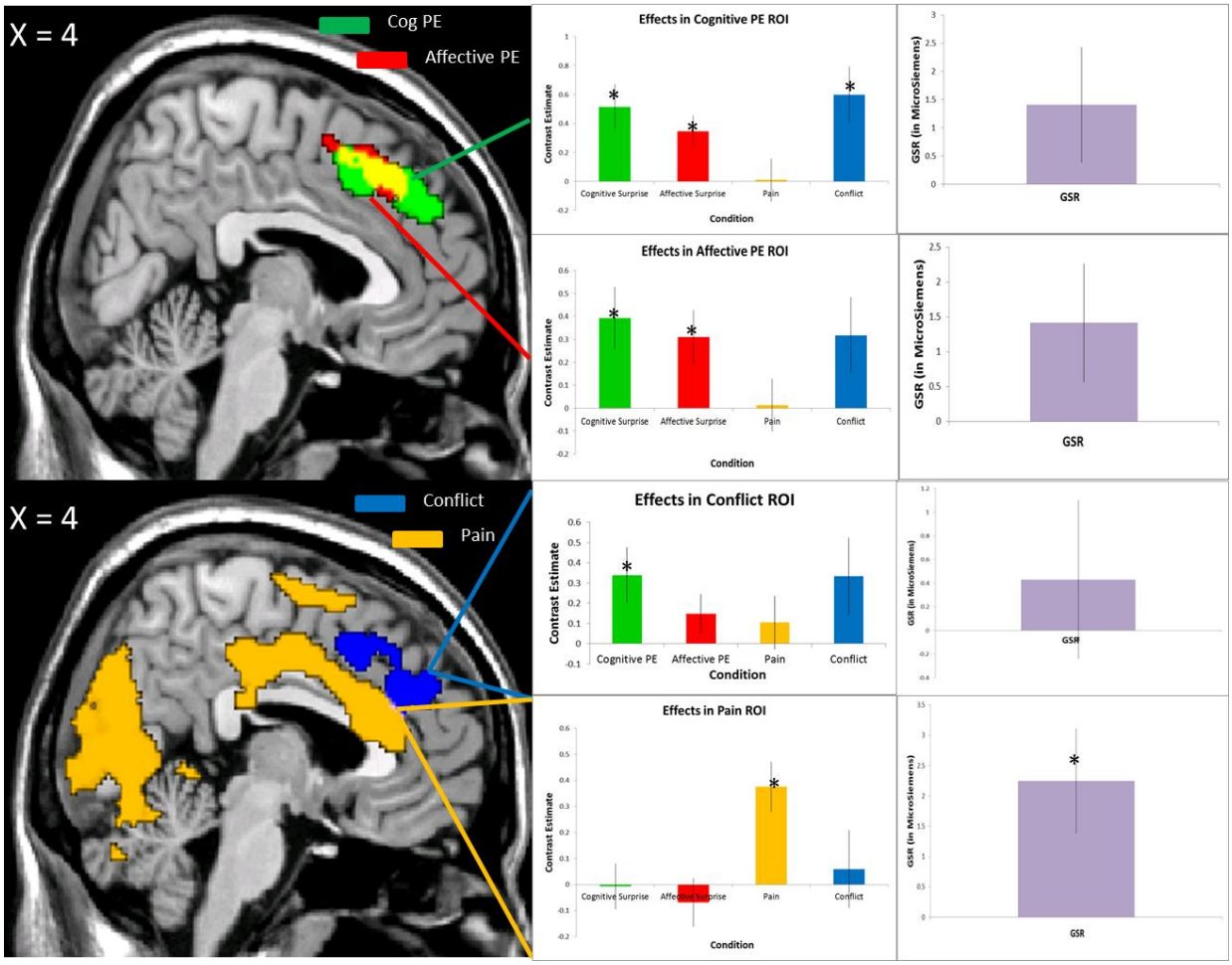

Figure 5: Cognitive PE, Affective PE, Conflict, and Pain ROIs defined by the leave-one-out cross-validation procedure (Esterman et al, 2010). Contrasts for each effect were extracted from each ROI separately, in addition to GSR (right column). All results depicted at a threshold of $p < 0.05$, cluster corrected. * = $p < 0.05$.

Morphological Influences on Prediction Error Effects

The finding that the pain and Stroop surprise effects were partially distinct but still somewhat overlapping raises the question of what accounts for the overlap. It could be that the representations are interdigitated in individual subjects, or it could be that the location of the pain and Stroop surprise effects varies across subjects, perhaps due to morphological differences. Recent studies have shown that morphological differences within the mPFC can affect the

location of activation profiles for certain psychological phenomena, such as error feedback (Amiez et al, 2013). In particular, subjects can show considerable variability in whether they have a paracingulate sulcus, an additional sulcus running dorsal to and roughly parallel with the cingulate sulcus (Misra et al, 2014). Although our group-level analyses showed effects of surprising outcomes in relatively homogenous areas, it is possible that subjects with a paracingulate sulcus may show activity in a different location than those subjects without a paracingulate sulcus. The resulting location of each effect may thus be the result of averaging across over these distinct activation locations.

To test this possibility, we categorized participants according to whether they had a paracingulate sulcus or not in either the left or right hemisphere. A subject was determined to have a paracingulate sulcus if they showed a sulcus running dorsal to the cingulate sulcus for at least 25mm and for at least 3 contiguous sagittal slices. A total of 18 subjects (78% of our sample) had at least one paracingulate sulcus, consistent with findings from prior studies (Paus et al, 1996). A total of 14 of these subjects had one in the right hemisphere, 9 subjects had one in the left hemisphere and 5 subjects had one in both the left and right hemispheres.

After dividing the subjects into right and left paracingulate sulcus groups, one-sample t-tests were carried out in each group for the effects of cognitive surprise, affective surprise, and the positive and negative prediction errors for both the pain and Stroop conditions as described in section 3.2.1. From this analysis, we did not observe any differences in the locations of pain and Stroop surprise effects as a function of whether subjects did or did not have a paracingulate sulcus. This suggests that the representations of pain and Stroop surprise may be partially interdigitated within subjects.

Chapter 5

Discussion

Study 1: Imagining Errors

These three neuroimaging studies provide some of the first empirical tests of the PRO model. The first sought to explore the neural mechanisms involved in imagining possible actions and predicting their potential consequences, a concept variously referred as mentation (Goldman-Rakic, 1996) or "dynamic evaluation lookahead" (van der Meer and Redish, 2010) based on learned R-O predictions (Colwill and Rescorla, 1990). We identified the mPFC as playing a potential role in action-outcome prediction. In prior studies, the mPFC has been implicated in predicting action outcomes (Brown and Braver, 2005; Valentin et al., 2007; Glascher et al., 2009; Krawitz et al., 2011) or similarly learning the value of actions (Kennerley et al., 2006; Rushworth et al., 2007), although previous studies have not isolated R-O prediction from the actual execution of the corresponding responses.

Because of the absence of explicit feedback and motor response during the Imagine condition, our findings in mPFC are unlikely to be accounted for by models assigning a role for error detection (Gehring et al., 1993). Our findings of greater ACC activity for imagining errors, combined with greater motor cortex activity representing the correct response while subjects imagined errors, might initially seem consistent with the response conflict model (Botvinick et al., 2001) as extended to anticipation (Sohn et al., 2007). Nevertheless, multiple responses can lead to ACC activity even without response conflict (Brown, 2009), which suggests that ACC may reflect the anticipated responses and outcomes rather than conflict *per se*. Furthermore, anticipatory effects in ACC likewise do not necessarily entail response conflict (Aarts et al., 2008). Another possible alternative account of ACC activity is that it correlates

with time on task (Grinband et al., 2010). We attempted to control for this by equalizing the duration period for imagining both correct and error outcomes. However, we cannot entirely rule out the possibility that participants spent unequal amounts of time imagining the correct vs. incorrect options.

Given the above, our results from this study are consistent with a comprehensive computational model of mPFC as anticipating and then evaluating the outcome of planned actions (Alexander & Brown, 2011). We have recently developed a new model of mPFC, the predicted response outcome (PRO) model, according to which R-O predictions are generated and subsequently evaluated against actual outcomes in the mPFC. A key prediction of the model is that mPFC (and especially ACC) signals a prediction of the anticipated outcome of an action, which may be subsequently compared against the actual outcome. In the model, more activity representing planned actions leads to greater activity in ACC representing anticipated outcomes (Brown, 2009), while discrepancies between actual and predicted action outcomes form the basis of the error effect in mPFC. These discrepancy signals are not limited to errors; they also signal surprisingly good outcomes (Jessup et al., 2010). There is ample evidence that surprising action outcomes are detected in part by ACC in monkeys (Ito et al., 2003; Hayden et al., 2011) and humans (Nee et al., 2011). Nevertheless, two theoretical questions remained open. First, it was unclear where the R-O predictions might originate from in humans, though at least stimulus if not action value may be represented in the OFC of humans (Valentin et al., 2007; Glascher et al., 2009), and actions may be simulated in the hippocampus (van der Meer and Redish, 2010). Second, it was unclear whether the R-O predictions would be represented in the mPFC even when action execution was not imminent. Our results are consistent with the PRO model predictions and indicate that mPFC activity may reflect a subjective prediction of action

outcomes. The present results suggest that these action-outcome predictions may account for the observed multiple response effects (Brown, 2009), which are driven as the actions are simulated in motor cortical areas. Our results further show that these signals are present even when action execution is merely imagined and not imminent. The region that responds to imagined errors overlaps with the region that responds to actual errors, which is consistent with a partial overlap between regions that predict outcomes and regions that evaluate actual outcomes. The finding of an effect of *imagined* error relative to correct outcomes, combined with the findings in motor cortex, suggest mechanisms by which the mPFC signal may effectively represent more of a subjective than objective outcome prediction. Nevertheless, we cannot completely exclude the possibility that mPFC activity reflects the mechanics of the planned response in addition to the anticipated outcome. These findings, combined with prior evidence that mPFC activity is key to risk avoidance (Brown and Braver, 2005; Magno et al., 2006; Brown and Braver, 2007), are consistent with proposals that mPFC is a region crucial to the ability to anticipate and avoid adverse consequences even when a risky action is not planned to be executed immediately. Indeed, overactivity of the mPFC and especially ACC appears to be a key ingredient in obsessive-compulsive disorder (Machlin et al., 1991), in which the excessive urge to avoid potential dangers may be experienced even when no action is otherwise imminent.

As a whole, the results are consistent with the PRO model account of the mPFC as involved in predicting the potential outcomes of an action. The mPFC results are consistent with a model in which mPFC evaluates potential outcomes with a view toward guiding decisions among possible actions, and our results show that this occurs even when action is not imminent. Our results provide a view of the networks involved in guiding decisions about actions and especially how those networks function when dissociated from action execution. These networks are

central to a number of clinical disorders, and a better understanding of their role is urgent given that the impaired ability to think about and take into account the outcomes or consequences of actions is a hallmark of various clinical disorders such as obsessive-compulsive compulsive disorder, schizophrenia, and drug abuse (Petry and Casarella, 1999; Bechara et al., 2002). The identification of the neural mechanisms involved in prospective decision-making has the potential to inform more effective pharmacological and cognitive treatments in patient populations.

Study 2: The PRO Model and Distinct Prediction and Outcome Evaluation Regions within the ACC

Next, using model-based regressors generated by the PRO model, we found that prediction-related signals loaded onto posterior and perigenual portions of the ACC. This prediction effect did not overlap with a medial supracollosal region of the ACC that showed a complementary effect of outcome evaluation (Figure 3). The finding of distinct prediction and evaluation regions within mPFC is consistent with the corresponding theoretically distinct prediction and evaluation mechanisms of the PRO model. Previous studies have shown anticipatory signals in ACC (Aarts et al., 2008; Sohn et al., 2007), but it has been unclear whether these areas of the ACC were the same as those that generate outcome-related signals (Dehaene et al., 1994; Gehring et al., 1993; Holroyd and Coles, 2002). Here, we found that the regions in mPFC encoding prediction signals are distinct from other regions of the mPFC that encode outcomes.

Other studies have suggested regional dissociations within the ACC in cognitive tasks, and our findings account for earlier regional dissociations in the framework of the PRO model. Behrens et al. (2007) reported distinct social vs. reward learning volatility effects in the anterior

cingulate gyrus and sulcus, respectively. We previously demonstrated that volatility effects could be understood as reflecting an outcome evaluation signal in the PRO model (Alexander & Brown, 2011). The current results show that the region with EVALUATION effects includes both of the regions where reward and social volatility effects have been reported (Behrens et al., 2007). Also, three distinct cingulate motor areas have recently been reported in the human (Amiez et al., 2012). Interestingly, the three human cingulate motor areas show substantial overlap with the corresponding three PREDICTION and EVALUATION regions found here in the ACC. The EVALUATION region in particular coincides with a region that has topographic connections with widespread regions of lateral prefrontal cortex (Beckmann et al., 2009; Blumenfeld et al., 2012; Taren et al., 2011).

Our results challenge other theories of ACC function, in particular conflict monitoring theory (Botvinick et al., 2001). Conflict monitoring theory posits that mutually incompatible response processes can account for greater activation within ACC. However, it is unclear how the conflict model could account for the observed PREDICTION regressor effects, because the time period of prediction is separate from the action periods when conflict might be present. Similarly, it is unclear how the conflict model could account for the EVALUATION regressor effects, as again there is no overt action associated with learning the outcome.

In contrast, the present results add to a growing body of findings that are consistent with the ACC as a region that predicts and evaluates outcomes, as exemplified in the PRO model. We have previously demonstrated that apparent conflict effects in ACC can be found even when the task is manipulated such that the responses are not in conflict with each other, although our previous design did not distinguish prediction vs. outcome signals (Brown, 2009). Nevertheless, such data cannot be accommodated by the conflict monitoring model, but according to the PRO

model, apparent conflict effects may instead result from a prediction of multiple possible responses on conflict trials (i.e. both correct and incorrect outcomes are possible with incongruent trials), vs. a single predicted outcome (i.e. correct on congruent trials). The present results now suggest that such a prediction signal may be presented in the posterior and perigenual ACC. Likewise, we and others have found that error effects (Gehring et al., 1993), which have been argued to represent conflict (Yeung et al., 2004), may instead represent surprise. In particular, error effects reverse when errors are more common than correct trials (Ferdinand et al., 2012; Jessup et al., 2010; Oliveira et al., 2007). The PRO model simulates both error and surprising correct outcome effects in a single EVALUATION regressor (Alexander and Brown, 2011), and the present results now suggest that such a signal is represented specifically in the mid-dorsal cingulate.

The results are consistent with a growing body of literature suggesting that ACC is involved in representing action values (Kennerley et al., 2006; Croxson et al., 2009; Glascher et al., 2009; Hayden et al., 2011). In addition, these results are consistent with other neuroimaging and modeling work of the mPFC, including Bayesian modeling of absolute prediction error between expectation and outcome phases (Ide et al., 2013), Bayesian modeling of hierarchical prediction errors (Iglesias et al., 2013), and updating one's prior beliefs about the environment in order to form more accurate predictions about response-outcome associations (O'Reilly et al., 2013). Notably, the prediction and outcome effects discussed in these studies show similar patterns of brain activity as shown in the current paper. Furthermore, the results provide empirical support for the theoretical prediction of two interacting prediction and evaluation components that subserve performance monitoring: predicted action values are represented in a network of regions including the ACC, and these in turn provide a basis against which other

regions of the ACC evaluate ongoing behavior. Actions that fail to yield an expected level of reward at the time of outcome may be evaluated within the ACC as requiring corrective action, such as a change in strategy (Hayden et al., 2011; Kennerley et al., 2006) or an impetus to forage in order to find a more valuable action (Kolling et al., 2012).

One potential issue with our model-based analysis is that outcomes presented at especially long jitter intervals are undersampled in our design, resulting in a failure of the PRO model to converge on appropriate predictions for outcomes presented following infrequent long intervals. In order to address this issue, we simulated the PRO model twice for each subject, once in order to generate parametric modulators for the PREDICTION regressor during which jittered intervals were simulated as they were experienced by the subject, and once in order to generate EVALUATION modulators, during which intervals were set to the most common jitter interval (60 model iterations). This approach ensured that model predictions at the time of feedback converged on the likelihood of observing the various outcomes associated with the task. Other possible strategies for addressing this issue are possible. One such strategy for generating parametric modulators might involve more extensive training of the model on synthetic data in order to resolve the problem with undersampling long-jitter trials in order to allow the model to converge on appropriate predictions. An additional option would be to model only those trials in our GLM with the most frequently observed jitter intervals. Finally, in order to prevent undersampling of specific intervals during the experiment, jitter intervals might be sampled from a uniform distribution of possible jitter intervals (rather than an exponential distribution), although this approach would impact overall efficiency of the experimental design (Dale, 1999).

Ruling Out Alternative Explanations

One potential explanation for the prediction results is that the dual-task nature of making two predictions could drive ACC activity. A previous study of dual-task performance found activation in perigenual ACC (Dreher and Grafman, 2003), and that region overlaps with the region found here in response to prediction. However, this previous study used a block design that did not distinguish response, prediction, and outcome feedback conditions as we have done here. Our results show that the ACC region with multiple outcome prediction effects is specifically active during the prediction phase of a trial as distinct from the response or outcome phase. Thus, is it likely that if the same region is active during dual-task performance, such activity may reflect predictions of the outcomes associated with performing each of the two tasks rather than task responses or feedback evaluation (Brown, 2009; Jahn et al., 2011).

One particularly interesting result of the EVALUATION analysis was a significant cluster of activation in the visual cortex, in addition to the observed dorsal ACC cluster. Surprising outcomes may call for increased attention to inputs reflected here in greater visual activity. By increasing the activity of inputs, the cognitive system may be better suited to gather contextual information that can account for discrepancies between expectations and outcomes thereby minimizing future prediction errors. While increased attention may explain both visual and ACC activation, it is unclear how such an account could explain the various effects ascribed to the ACC that are predicted by the PRO model. Instead, we suggest that distinct mechanisms govern the ACC and visual activations.

Another set of confounding factors to be ruled out is the potential effect of errors or error likelihood. In the *Predict2* condition, there were two opportunities to fail at finding the stay cue, which in principle might lead to greater error likelihood effects in the *Predict2* vs. *Predict1*

conditions, as well as potentially greater error effects. However, we designed the task to dissociate errors vs. the absence of one or both stay cues. Subjects were given a monetary incentive to perform the task correctly. Crucially, the reward was given for following the win-stay/lose-shift strategy, and this contingency was explained explicitly to the participants as part of the task instructions. The reward available did not differ between *Predict1* vs. *Predict2* trials, nor was the reward reduced if subjects received a switch cue, provided that they followed the task rules. In this way, even though receiving a switch cue was unexpected, it was not to be considered an error provided that subjects followed the task rules. Thus, the effects of one or two switch cues can be attributed to surprise or switching, but not to errors in terms of gaining reward. Furthermore, the error rates were low overall, so the effects are not likely to be attributable to differences in error likelihood across conditions. If anything, participants were more likely to commit error in the *Predict1* condition compared to the *Predict2* condition, which would argue against an interpretation of this effect in terms of error likelihood. Thus, it is unlikely that the observed effects represent error-related processes.

Although several other computational models of mPFC function could be considered, the task design makes it difficult to carry out a quantitative model comparison. As the current paradigm modeled both prediction and outcome phases, it would not be a direct comparison to include models for which model behavior is undefined for prediction (e.g., reinforcement learning; Holroyd & Coles, 2002), undefined for outcome (e.g., error likelihood; Brown & Braver, 2005), or undefined for both prediction and outcome (e.g., conflict monitoring or time on task; Botvinick et al, 2004; Grinband et al, 2010). Furthermore, a model of mPFC activity such as the reward value and prediction model (RVPM; Silvetti et al, 2011) is too similar to the model used here to serve as a viable alternative model.

Overall, it is unclear how existing theories other than the PRO model could account for the present results. Other proposed theories cast ACC as computing error likelihood (Brown and Braver, 2005), volatility (Behrens et al., 2007), time-on-task (Carp et al., 2010; Grinband et al., 2011), differences between actual vs. intended responses (Scheffers and Coles, 2000), differences between actual vs. intended outcomes (Holroyd and Coles, 2002; Ito et al., 2003), and predicted action values (Scheffers and Coles, 2000; Walton et al., 2004). The prediction effect occurs at a time that is temporally dissociated from response processes, so it is unlikely to involve response conflict. The interval between prediction and outcome has the same distribution in the *Predict1* and *Predict2* conditions, so it is unclear how a time-on-task account could explain the prediction effect. Furthermore, the outcome events are modeled separately from the prediction events and with a variable interval between them, so the prediction and outcome events can be estimated independently of each other. Lastly, the nature of the task contingencies does not change throughout the course of the experiment, so volatility differences are unlikely to play a role. To the best of our knowledge, the PRO model is the only existing framework that can account for the multiple outcome prediction effect found here. Furthermore, we have recently shown that a computational simulation of the PRO model can reproduce the various effects that have been cited as evidence for all of these various theories of ACC function above (Alexander and Brown, 2011), as well as generating the regressors used to model the prediction and outcome effects identified here. Thus, our results are consistent with the PRO model as a unifying theory of ACC function.

Study 3: Multimodal representation of prediction error within the mPFC

Several recent studies have highlighted the role of the dACC and SMA in processing surprise, especially negative surprise, i.e. the omission of an expected outcome (Alexander & Brown, 2011). We found activation for surprising outcomes, regardless of the valence of those outcomes, and these were not accounted for by physiological measures such as GSR.

In particular, we observed a significant overlap of activation within the SMA/dACC region for several different types of surprise. Most notably, both better-than-expected and worse-than-expected pain outcomes overlapped (but only partially) within this region, suggesting that surprise contributes to the observed effects, and in a way that is distinct from physiological arousal effects. Logically, because pain and Stroop surprise activate different regions under different conditions, they cannot both be driven solely by a single physiological arousal signal. Likewise regarding Stroop outcomes, both pPE and nPE outcomes gave rise to increased mPFC activity, but in distinct cortical areas: pPE outcomes were found primarily within the SMA region and overlapped with the observed nPE and pPE pain outcomes, while nPE Stroop outcomes were observed in the rostral anterior cingulate.

The overlap between the pPE Stroop outcome and the pPE and nPE pain outcomes could be explained by the negative surprise component of the PRO model (Alexander & Brown, 2010; 2011). According to the PRO model, any expected event in the environment that fails to occur will elicit a surprise signal. In the case of a cue that predicts the occurrence of an incongruent stimulus, the presentation of a congruent stimulus entails the omission of the expected incongruent stimulus, and this omission constitutes a "negative" surprise: the unexpected non-occurrence of the incongruent stimulus. It is important to note that negative surprise here refers

to omission, which is independent of valence and may be subjectively better or worse than expected.

Comparing the PRO Model to Reinforcement Learning Models of mPFC Activity

Within the framework of reinforcement learning theory (Holroyd & Coles, 2002), negatively valenced events should elicit more activity in the mPFC than positively valenced events. Therefore, in our current paradigm in which we matched for the unexpectedness of both pPE and nPE outcomes, reinforcement learning theory predicts that activity in ACC should be significantly greater for nPE outcomes than for pPE outcomes. Within the context of the PRO model, on the other hand (Alexander & Brown, 2011), the primary driver behind mPFC activity when evaluating outcomes is the unexpectedness of the outcome, regardless of the valence. Our results, showing activation for both pPE and nPE outcomes regardless of valence, are more consistent with the PRO model framework of mPFC activity.

The current study speaks to several contemporary neuroimaging studies of mPFC function and modeling work as well. Roy et al. (2009) found signals reflecting the surprisingness of pain, but their results are slightly more ventral to our findings (possibly reflecting the administration of electrical shock to the ankle, a different anatomical site than the one used in this study), and they did not directly compare affective vs. cognitive PE, as we have here. Several recent EEG studies (Ferdinand et al., 2012; Oliveira et al., 2007) showed that both negative and positive outcomes matched for infrequency (a concept similar to the one of unexpectedness used here) elicited similar deflections across electrodes placed on frontal scalp electrodes approximately above the dACC/SMA area, consistent with our present results and the PRO model framework. The current study provides converging evidence to suggest that the

deflections measured by these electrodes are located specifically in the dACC/SMA zone. Similarly, a study carried out by Bonini et al (2014) using intracranial EEG found that the SMA appears to be involved in the early evaluation of the outcome of actions, in case these actions need to be corrected as they are being initiated, or corrected on a future trial.

In addition, this study complements fMRI studies examining the effects of both positive and negative surprise in the reward domain in the mPFC. In one such study by Jessup et al (2010), the mPFC was responsive not only to infrequent negative outcomes, but also to infrequent positive outcomes, analogous to winning a lottery where the winnings are desirable but unlikely to occur. Furthermore, a study analyzing the mPFC using regressors generated by the PRO model found anatomically distinct regions of the mPFC involved in both prediction and outcome evaluation (Jahn et al, 2014). In that study, the PRO model prediction layer signals correlated with activity in the caudal cingulate zone and rostral ACC, while outcome layer signals correlated with activity in the dACC. The surprise effects predicted by the PRO model were found in a region approximately overlapping with the multi-modal surprise (prediction error) effects found here (Jahn et al, 2014).

A further analysis of prediction-related effects within several different ROIs within the ACC revealed no significant effects for cues predicting either pain or Stroop outcomes, similar to results reported in Aarts et al (2011) in which congruent and incongruent spatial Stroop stimuli were cued probabilistically. Within the framework of the PRO model, effects related to event prediction are expected to attenuate as the events to be predicted become less consistent. Thus, while the PRO model predicts increased activity in ACC associated with cues which indicate likely congruent or incongruent trials, previous studies, as well as the present one, are likely inadequately powered to uncover these effects.

Summary of Findings

Taken together, these findings suggest that the mPFC region is responsive to the unexpectedness of an event, as opposed to merely the valence of the outcome, and that this function is general across various modalities that are nonetheless represented distinctly in the medial wall. The findings of these studies also suggest that across several different experimental paradigms, both the conflict monitoring and reinforcement learning models, and that the PRO model can be used to simulate both prediction- and outcome-related activity in the mPFC.

The final study was designed to build upon the previous two studies by testing whether the mPFC acts as a modality-general processor of prediction error. In addition, the inclusion of both positively- and negatively-valenced stimuli allowed for examining whether prediction error is modulated by the valence of the outcome, as predicted by the reinforcement learning model, or whether it is processed by the mPFC independently of valence, as predicted by the PRO model.

In this study we combined different modalities of surprise, specifically pain and Stroop stimuli, and found that many of these effects partially overlapped with each other in the dACC/SMA. This complements other studies showing monetary reward prediction error effects in overlapping regions (Jessup et al., 2010). Our findings cannot be readily explained by competing models of mPFC activity, such as the conflict monitoring model (Botvinick et al, 2002) or classical reinforcement learning models (Holryod & Coles, 2002). First, there was no overt conflict in the pain condition, as no response was required when the outcome was received. Second, we found effects of the surprising absence of response conflict, namely in the pPE Stroop condition.

Overall, our findings across all three studies are consistent with the PRO model of mPFC activity, and are consistent with several other findings detailing this region's involvement in

processing both positive and negative surprise. This provides a solid foundation for future empirical studies of the mPFC, and explains several of the effects observed there within a framework that casts the mPFC as an action-outcome predictor.

References

Aarts E, Roelofs A (2011) Attentional control in anterior cingulate cortex based on probabilistic cueing. J Cogn Neurosci 23:716-727.

Aarts E, Roelofs A, van Turennout M (2008) Anticipatory activity in anterior cingulate cortex can be independent of conflict and error likelihood. J Neurosci 28:4671-4678.

Adolphs, R., Tranel, D., & Damasio, A. R. (2003). Dissociable neural systems for recognizing emotions. Brain and cognition, 52(1), 61-69.

Ahmed, S. H., & Koob, G. F. (2005). Transition to drug addiction: a negative reinforcement model based on an allostatic decrease in reward function. Psychopharmacology, 180(3), 473-490.

Alexander W, Brown J (2010) Computational models of performance monitoring and cognitive control. Topics in Cognitive Science 2:658-677.

Alexander, W.H., Brown, J.W., 2011. Medial prefrontal cortex as an action-outcome predictor. Nat Neurosci 14, 1338–1344.

Alexander, W.H., Fukunaga, R., Finn, P, & Brown, J. (in press). Reward salience and risk aversion underlie differential ACC activity in substance dependence. To appear in: Neuroimage: Clinical.

Allman, J. M., Hakeem, A., Erwin, J. M., Nimchinsky, E., & Hof, P. (2001). The anterior cingulate cortex. Annals of the New York Academy of Sciences, 935(1), 107-117.

Amador, N., Schlag-Rey, M., & Schlag, J. (2004). Primate antisaccade. II. Supplementary eye field neuronal activity predicts correct performance. Journal of neurophysiology, 91(4), 1672-1689

Amador, N., Schlag-Rey, M., Schlag, J., 2000. Reward-predicting and reward-detecting neuronal activity in the primate supplementary eye field. J. Neurophysiol. 84, 2166–70.

Amiez, C., Joseph, J. P., & Procyk, E. (2006). Reward encoding in the monkey anterior cingulate cortex. Cerebral Cortex, 16(7), 1040-1055.

Amiez, C., Neveu, R., Warrot, D., Petrides, M., Knoblauch, K., Procyk, E., 2013. The Location of Feedback-Related Activity in the Midcingulate Cortex Is Predicted by Local Morphology. J. Neurosci. 33, 2217–2228.

Amiez, C., Petrides, M., (2012). Neuroimaging Evidence of the Anatomo-Functional Organization of the Human Cingulate Motor Areas. Cereb. Cortex 48, 46–57.

Badgaiyan, R. D., & Posner, M. I. (1998). Mapping the cingulate cortex in response selection and monitoring. Neuroimage, 7(3), 255-260.

Baliki, M. N., Geha, P. Y., & Apkarian, A. V. (2009). Parsing pain perception between nociceptive representation and magnitude estimation. Journal of Neurophysiology, 101(2), 875-887.

Bayer, H. M., & Glimcher, P.W. (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron, 47(1), 129-141.

Bechara A, Dolan S, Hindes A (2002) Decision-making and addiction (part II): myopia for the future or hypersensitivity to reward? Neuropsychologia 40:1690-1705.

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. Cognition, 50(1), 7-15.

Beckmann, M., Johansen-Berg, H., Rushworth, M.F.S., 2009. Connectivity-based parcellation of human cingulate cortex and its relation to functional specialization. J. Neurosci. 29, 1175–90.

Behrens, T.E., Woolrich, M.W., Walton, M.E., Rushworth, M.F., 2007. Learning the value of information in an uncertain world. Nat. Neurosci. 10, 1214–1221.

Blumenfeld, R.S., Parks, C.M., Yonelinas, A.P., Ranganath, C., 2012. Putting the Pieces Together: The Role of Dorsolateral Prefrontal Cortex in Relational Memory Encoding. J. Cogn. Neurosci. 23, 257–265.

Bolla, K. I., Eldreth, D. A., London, E. D., Kiehl, K. A., Mouratidis, M., Contoreggi, C., & Ernst, M. (2003). Orbitofrontal cortex dysfunction in abstinent cocaine abusers performing a decision-making task. Neuroimage, 19(3), 1085-1094.

Botvinick MM, Braver TS, Barch DM, Carter CS, Cohen JD (2001) Conflict monitoring and cognitive control. Psychol Rev 108:624-652.

Botvinick, M. M. (2007). Conflict monitoring and decision making: reconciling two perspectives on anterior cingulate function. Cognitive, Affective, & Behavioral Neuroscience, 7(4), 356-366.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S. & Cohen, J. D. (2002). Conflict monitoring and cognitive control. Psychological Review, 108, 624–652.

Botvinick, M. M., Nystrom, L., Fissel, K., Carter, C.S., Cohen, J.D. (1999). Conflict monitoring versus selection-for-action in anterior cingulate cortex. Nature, 402, 179-181.

Botvinick, M.M., Cohen, J.D., Carter, C.S., 2004. Conflict monitoring and anterior cingulate cortex: an update. Trends Cogn. Sci. 8, 539–46.

Braver, T. S., Barch, D. M., & Cohen, J. D. (1999). Cognition and control in schizophrenia: a computational model of dopamine and prefrontal function. Biological Psychiatry, 46(3), 312-328.

Brown JW (2009) Conflict effects without conflict in anterior cingulate cortex: multiple response effects and context specific representations. Neuroimage 47:334-341.

Brown JW, Braver TS (2005) Learned Predictions of Error Likelihood in the Anterior Cingulate Cortex. Science 307:1118-1121.

Brown, J. W., & Braver, T. S. (2007). Risk prediction and aversion by anterior cingulate cortex. Cognitive, Affective, and Behavioral Neuroscience, 7(4), 266-277.

Brown, J. W., & Braver, T. S. (2008). A computational model of risk, conflict, and individual difference effects in the anterior cingulate cortex. Brain Research, 1202, 99-108.

Brown, J.W., 2011. Medial prefrontal cortex activity correlates with time-on-task: what does this tell us about theories of cognitive control? Neuroimage 57, 314–5.

Bush, G., Luu, P., Posner, M., 2000. Cognitive and emotional influences in anterior cingulate cortex. Trends Cogn. Sci. 4, 215–222.

Calder, A. J., Keane, J., Manes, F., Antoun, N., & Young, A. W. (2000). Impaired recognition and experience of disgust following brain injury. Nature neuroscience, 3(11), 1077-1078.

Carp, J., Kim, K., Taylor, S. F., Fitzgerald, K. D., & Weissman, D. H. (2010). Conditional differences in mean reaction time explain effects of response congruency, but not accuracy, on posterior medial frontal cortex activity. Frontiers in human neuroscience, 4.

Carp, J., Kim, K., Taylor, S.F., Fitzgerald, K.D., Weissman, D.H., 2010. Conditional Differences in Mean Reaction Time Explain Effects of Response Congruency, but not Accuracy, on Posterior Medial Frontal Cortex Activity. Front. Hum. Neurosci. 4, 231.

Carter, C. S., Botvinick, M. M., & Cohen, J. D. (1999). The contribution of the anterior cingulate cortex to executive processes in cognition. Reviews in the Neurosciences, 10(1), 49.

Carter, C.S., Braver, T.S., Barch, D.M., Botvinick, M.M., Noll, D., Cohen, J.D., 1998. Anterior cingulate cortex, error detection, and the online monitoring of performance. Science (80-.). 280, 747.

Carter, C.S., Braver, T.S., Barch, D.M., Botvinick, M.M., Noll, D., Cohen, J.D., 1998. Anterior cingulate cortex, error detection, and the online monitoring of performance. Science 280, 747-749.

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: the BIS/BAS scales. Journal of personality and social psychology, 67(2), 319.

Chandrasekhar, P.V.S., Capra, C.M., Moore, S., Noussair, C., Berns, G.S., 2008. Neurobiological regret and rejoice functions for aversive outcomes. Neuroimage 39, 1472–84.

Cole, M. W., Yeung, N., Freiwald, W. A., & Botvinick, M. (2009). Cingulate cortex: diverging data from humans and monkeys. Trends in neurosciences, 32(11), 566-574.

Cole, M.W., Yeung, N., Freiwald, W. a, Botvinick, M., 2009. Cingulate cortex: diverging data from humans and monkeys. Trends Neurosci. 32, 566–74.

Colwill R, Rescorla R (1990) Evidence for the hierarchical structure of instrumental learning. Animal Learning & Behavior 18:71-82.

Craig, A. D. (2002). How do you feel? Interoception: the sense of the physiological condition of the body. Nature Reviews Neuroscience, 3(8), 655-666.

Craig, A. D. (2003). Interoception: the sense of the physiological condition of the body. Current opinion in neurobiology, 13(4), 500-505.

Critchley, H. D., Wiens, S., Rotshtein, P., Öhman, A., & Dolan, R. J. (2004). Neural systems supporting interoceptive awareness. Nature neuroscience, 7(2), 189-195.

Croxson, P.L., Walton, M.E., O'Reilly, J.X., Behrens, T.E.J., Rushworth, M.F.S., 2009. Effort-based cost-benefit valuation and the human brain. J. Neurosci. 29, 4531–41.

Cunningham, W. A., Arbuckle, N. L., Jahn, A., Mowrer, S. M., & Abduljalil, A. M. (2010). Aspects of neuroticism and the amygdala: chronic tuning from motivational styles. Neuropsychologia, 48(12), 3399-3404.

Dale AM (1999) Optimal experimental design for event-related fMRI. Human Brain Mapping 8:109-114.

Dale, a M., 1999. Optimal experimental design for event-related fMRI. Hum. Brain Mapp. 8, 109–14.

Daw ND, Niv Y, Dayan P (2005) Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat Neurosci 8:1704-1711.

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. Nature, 441(7095), 876-879.

Dawe, S., Gullo, M. J., & Loxton, N. J. (2004). Reward drive and rash impulsiveness as dimensions of impulsivity: implications for substance misuse. Addictive behaviors, 29(7), 1389-1405.

Dayan P, Niv Y (2008) Reinforcement learning: the good, the bad and the ugly. Curr Opin Neurobiol 18:185-196.

de Mendizábal, N. V., Jones, D. R., Jahn, A., Bies, R. R., & Brown, J. W. (2014). Nicotine and Cotinine Exposure from Electronic Cigarettes: A Population Approach. Clinical pharmacokinetics, 1-12.

Dehaene, S., Posner, M. I., & Tucker, D. M. (1994). Localization of a neural system for error detection and compensation. Psychological Science, 303-305.

Deichmann R, Gottfried JA, Hutton C, Turner R (2003) Optimized EPI for fMRI studies of the orbitofrontal cortex. Neuroimage 19:430-441.

Deichmann, R., Gottfried, J.A., Hutton, C., Turner, R., 2003. Optimized EPI for fMRI studies of the orbitofrontal cortex. Neuroimage 19, 430–441.

Derbyshire, S., Vogt, B., Jones, A., 1998. Pain and Stroop interference tasks activate separate processing modules in anterior cingulate cortex. Exp. Brain Res. 118, 52–60.

Di Chiara, G., Bassareo, V., Fenu, S., De Luca, M. A., Spina, L., Cadoni, C., & Lecca, D. (2004). Dopamine and drug addiction: the nucleus accumbens shell connection. Neuropharmacology, 47, 227-241.

Drabant, E.M., Kuo, J.R., Ramel, W., Blechert, J., Edge, M.D., Cooper, J.R., Goldin, P.R., Hariri, A.R., Gross, J.J., 2011. Experiential, autonomic, and neural responses during threat anticipation vary as a function of threat intensity and neuroticism. Neuroimage 55, 401–10.

Dreher, J. C., & Grafman, J. (2003). Dissociating the roles of the rostral anterior cingulate and the lateral prefrontal cortices in performing two tasks simultaneously or successively. Cerebral cortex, 13(4), 329-339.

Egner, T., Monti, J.M., Summerfield, C., 2010. Expectation and surprise determine neural population responses in the ventral visual stream. J. Neurosci. 30, 16601–8.

Etkin, A., Egner, T., Kalisch, R., 2011. Emotional processing in anterior cingulate and medial prefrontal cortex. Trends Cogn. Sci. 15, 85–93.

Falkenstein, M., Hohnsbein, J., Hoormann, J., Blanke, L., 1991. Effects of crossmodal divided attention on late ERP components. II. Error processing in choice reaction tasks. Electroencephalogr. Clin. Neurophysiol. 78, 447–55.

Fan, J., Hof, P.R., Guise, K.G., Fossella, J. a, Posner, M.I., 2008. The functional integration of the anterior cingulate cortex during conflict processing. Cereb. Cortex 18, 796–805.

Fellows, L.K., Farah, M.J., 2005. Is anterior cingulate cortex necessary for cognitive control? Brain 128, 788–796.

Ferdinand, N. K., Mecklinger, A., Kray, J., & Gehring, W. J. (2012). The processing of unexpected positive response outcomes in the mediofrontal cortex. The Journal of Neuroscience, 32(35), 12087-12092.

Ferdinand, N.K., Mecklinger, A., Kray, J., Gehring, W.J., 2012. The Processing of Unexpected Positive Response Outcomes in the Mediofrontal Cortex 32, 12087–12092.

Franken, I. H. (2002). Behavioral approach system (BAS) sensitivity predicts alcohol craving. Personality and Individual Differences, 32(2), 349-355.

Franken, I. H., & Muris, P. (2006). BIS/BAS personality characteristics and college students' substance use. Personality and Individual Differences, 40(7), 1497-1503.

Franken, I. H., Rassin, E., & Muris, P. (2007). The assessment of anhedonia in clinical and non- Journal of affective disorders, 99(1), 83-89.

Garavan, H., Ross, T. J., Murphy, K., Roche, R. A. P., & Stein, E. A. (2002). Dissociable executive functions in the dynamic control of behavior: inhibition, error detection, and correction. Neuroimage, 17(4), 1820-1829.

Garofalo, S., Maier, M.E., di Pellegrino, G., 2014. Mediofrontal negativity signals unexpected omission of aversive events. Sci. Rep. 4, 4816.

Gehring WJ, Coles MGH, Meyer DE, Donchin E (1990) The error-related negativity: An event-related potential accompanying errors. Psychophysiology 27:S34.

Gehring WJ, Goss B, Coles MGH, Meyer DE, Donchin E (1993) A Neural System for Error-Detection and Compensation. Psychol Sci 4:385-390.

Gehring, W. J., & Willoughby, A. R. (2002). The medial frontal cortex and the rapid processing of monetary gains and losses. Science, 295(5563), 2279-2282.

Gehring, W. J., Knight, R. T. (2000). Prefrontal-cingulate interactions in action monitoring. Nature Neuroscience, 3, 516-520.

Gehring, W.J., Goss, B., Coles, M.G.H., Meyer, D.E., Donchin, E., 1993. A neural system for error detection and compensation. Psychol. Sci. 4, 385–390.

Gehring, W.J., Goss, B., Coles, M.G.H., Meyer, D.E., Donchin, E., 1993. A Neural System for Error-Detection and Compensation. Psychol. Sci. 4, 385–390.

Gemba, H., Sasaki, K., & Brooks, V. B. (1986). 'Error'potentials in limbic cortex (anterior cingulate area 24) of monkeys during motor learning. Neuroscience letters, 70(2), 223-227

Glascher J, Daw N, Dayan P, O'Doherty JP (2010) States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. Neuron 66:585-595.

Glascher J, Hampton AN, O'Doherty JP (2009) Determining a role for ventromedial prefrontal cortex in encoding action-based value signals during reward-related decision making. Cereb Cortex 19:483-495.

Gläscher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. Neuron, 66(4), 585-595.

Gläscher, J., Hampton, A.N., O'Doherty, J.P., 2009. Determining a role for ventromedial prefrontal cortex in encoding action-based value signals during reward-related decision making. Cereb. Cortex 19, 483–95.

Glimcher, P.W. & Rustichini, A. (2004). Neuroeconomics: The consilience of brain and

Goldman-Rakic PS (1996) The prefrontal landscape: Implications of functional architecture for understanding human mentation and the central executive. In: The prefrontal cortex: Executive and cognitive functions (Roberts AC, Robbins TW, Weiskrantz L, eds), pp 87-103. Oxford: Oxford University Press.

Gray, J. A. (1970). The psychophysiological basis of introversion-extraversion. Behavior Research and Therapy, 8, 249-266.

Grinband J, Savitskaya J, Wager TD, Teichert T, Ferrera VP, Hirsch J (2010) The dorsal medial frontal cortex is sensitive to time on task, not response conflict or error likelihood. Neuroimage, 57(2), 320-322.

Grinband, J., Savitskaya, J., Wager, T. D., Teichert, T., Ferrera, V. P., & Hirsch, J. (2011). Conflict, error likelihood, and RT: Response to Brown & Yeung et al. NeuroImage,

Hadland, K. A., Rushworth, M. F., Gaffan, D. & Passingham, R. E. (2003). The anterior

Hassabis D, Kumaran D, Vann SD, Maguire EA (2007) Patients with hippocampal amnesia cannot imagine new experiences. Proc Natl Acad Sci U S A 104:1726-1731.

Hayden BY, Heilbronner SR, Pearson JM, Platt ML (2011) Surprise signals in anterior cingulate cortex: neuronal encoding of unsigned reward prediction errors driving adjustment in behavior. J Neurosci 31:4178-4187.

Hayden, B. Y., Pearson, J. M., & Platt, M. L. (2011). Neuronal basis of sequential foraging decisions in a patchy environment. Nature neuroscience, 14(7), 933-939.

Hayden, B.Y., Heilbronner, S.R., Pearson, J.M., Platt, M.L., 2011. Surprise signals in anterior cingulate cortex: neuronal encoding of unsigned reward prediction errors driving adjustment in behavior. J. Neurosci. 31, 4178–4187.

Hayden, B.Y., Pearson, J.M., Platt, M.L., 2009. Fictive reward signals in the anterior cingulate cortex. Science 324, 948–50.

Hester, R., Fassbender, C., & Garavan, H. (2004). Individual differences in error processing: a review and reanalysis of three event-related fMRI studies using the GO/NOGO task. Cerebral Cortex, 14(9), 986-994.

Hochman, E. Y., Wang, S., Milner, T. E., & Fellows, L. K. (2015). Double dissociation of error inhibition and correction deficits after basal ganglia or dorsomedial frontal damage in humans. Neuropsychologia.

Hohnsbein J, Falkenstein M, Hoorman J (1989) Error processing in visual and auditory choice reaction tasks. Journal of Psychophysiology 3:32.

Holroyd, C.B., Coles, M.G., 2002. The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. Psych. Rev. 109, 679–709.

Holroyd, C.B., Coles, M.G.H., 2002. The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. Psychol. Rev. 109, 679–709.

Ide, J.S., Shenoy, P., Yu, A.J., Li, C.R., 2013. Bayesian Prediction and Evaluation in the Anterior Cingulate Cortex 33, 2039–2047.

Iglesias, S., Mathys, C., Brodersen, K.H., Kasper, L., Piccirelli, M., den Ouden, H.E.M., Stephan, K.E., 2013. Hierarchical Prediction Errors in Midbrain and Basal Forebrain during Sensory Learning. Neuron 80, 519–530.

Isoda, M., & Hikosaka, O. (2007). Switching from automatic to controlled action by monkey medial frontal cortex. Nature neuroscience, 10(2), 240-248.

Ito S, Stuphorn V, Brown J, Schall JD (2003) Performance Monitoring by Anterior Cingulate Cortex During Saccade Countermanding. Science 302:120-122.

Ito, S., Stuphorn, V., Brown, J. W., & Schall, J. D. (2003). Performance monitoring by the anterior cingulate cortex during saccade countermanding. Science, 302(5642), 120-122.

Ito, S., Stuphorn, V., Brown, J.W., Schall, J.D., 2003. Performance monitoring by the anterior cingulate cortex during saccade countermanding. Science (80-.). 302, 120–122.

Jabbi, M., Bastiaansen, J., & Keysers, C. (2008). A common anterior insula representation of disgust observation, experience and imagination shows divergent functional connectivity pathways. PloS one, 3(8), e2939.

Jahn, A., Nee, D. E., Alexander, W. H., & Brown, J. W. (2014). Distinct regions of anterior cingulate cortex signal prediction and outcome evaluation. Neuroimage, 95, 80-89.

Jahn, A., Nee, D.E., Brown, J.W., 2011. The neural basis of predicting the outcomes of imagined actions. Front. Neurosci. 5, 128.

Jahn, A., Nee, D.E., Brown, J.W., 2011. The neural basis of predicting the outcomes of imagined actions. Front. Neurosci. 5, 128.

Jessup RK, Busemeyer JR, Brown JW (2010) Error effects in anterior cingulate cortex reverse when error likelihood is high. J Neurosci 30:3467-3472.

Jessup, R.K., Busemeyer, J.R., Brown, J.W., 2010. Error effects in anterior cingulate cortex reverse when error likelihood is high. J. Neurosci. 30, 3467–3472.

Jessup, R.K., Busemeyer, J.R., Brown, J.W., 2010. Error effects in anterior cingulate cortex reverse when error likelihood is high. J. Neurosci. 30, 3467–3472.

Johnson SH (2000) Thinking ahead: the case for motor imagery in prospective judgements of prehension. Cognition 74:33-70.

Johnson, M. K., Raye, C. L., Mitchell, K. J., Touryan, S. R., Greene, E. J., & Nolen-Hoeksema, S. (2006). Dissociating medial frontal and posterior cingulate activity during self-reflection. Social cognitive and affective neuroscience, 1(1), 56-64.

Kaelbling LP, Littman ML, Moore AW (1996) Reinforcement learning: A survey. Journal of Artificial Intelligence Research 4:237-285.

Kennerley SW, Walton ME, Behrens TE, Buckley MJ, Rushworth MF (2006) Optimal decision making and the anterior cingulate cortex. Nature Neuroscience 9:940-947.

Kennerley, S. W., & Wallis, J. D. (2009). Evaluating choices by single neurons in the frontal lobe: outcome value encoded across multiple decision variables. European Journal of Neuroscience, 29(10), 2061-2073.

Kennerley, S. W., Behrens, T. E., & Wallis, J. D. (2011). Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. Nature neuroscience, 14(12), 1581-1589.

Kennerley, S.W., Walton, M.E., Behrens, T.E., Buckley, M.J., Rushworth, M.F., 2006. Optimal decision making and the anterior cingulate cortex. Nat Neurosci 9, 940–947.

Kennerley, S.W., Walton, M.E., Behrens, T.E., Buckley, M.J., Rushworth, M.F., 2006. Optimal decision making and the anterior cingulate cortex. Nat Neurosci 9, 940–947.

Kerns, J.G., Cohen, J.D., MacDonald 3rd, A.W., Cho, R.Y., Stenger, V.A., Carter, C.S., MacDonald, A.W., 2004. Anterior Cingulate conflict monitoring and adjustments in control. Science 303, 1023–1026.

Kluver, H., & Bucy, P.C. (1939). Preliminary analysis of functions of the temporal lobes in monkeys. Archives of Neurological Psychiatry, 42, 979-1000.

Knutson, B., Westdorp, A., Kaiser, E., & Hommer, D. (2000). FMRI visualization of brain activity during a monetary incentive delay task. Neuroimage, 12(1), 20-27.

Kolling, N., Behrens, T.E.J., Mars, R.B., Rushworth, M.F.S., 2012. Neural mechanisms of foraging. Science (80-.). 336, 95–8.

Koob, G. F. (1999). The role of the striatopallidal and extended amygdala systems in drug addiction. Annals of the New York Academy of Sciences, 877(1), 445-460.

Koob, G. F., & Le Moal, M. (2005). Plasticity of reward neurocircuitry and the'dark side'of drug addiction. Nature neuroscience, 8(11), 1442-1444.

Kornhuber HH, Deecke L (1965) Hirnpotentialänderungen bei Willkürbewegungen und passiven Bewegungen des Menschen: Bereitschaftspotential und reafferente Potentiale (Changes in brain potentials with willful and passive movements in humans: the readiness potential and reafferent potentials). Pflügers Arch 284:1-17.

Krawitz A, Braver TS, Barch DM, Brown JW (2011) Impaired error-likelihood prediction in medial prefrontal cortex in schizophrenia. Neuroimage 54:1506-1517.

Krawitz, A., Braver, T.S., Barch, D.M., Brown, J.W., 2011. Impaired error-likelihood prediction in medial prefrontal cortex in schizophrenia. Neuroimage 54, 1506–1517.

Krawitz, A., Braver, T.S., Barch, D.M., Brown, J.W., 2011. Impaired error-likelihood prediction in medial prefrontal cortex in schizophrenia. Neuroimage 54, 1506–17.

Krawitz, A., Fukunaga, R., & Brown, J. W. (2010). Anterior insula activity predicts the influence of positively framed messages on decision making. Cognitive, Affective, & Behavioral Neuroscience, 10(3), 392-405.

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S., & Baker, C. I. (2009). Circular analysis in systems neuroscience: the dangers of double dipping. Nature neuroscience, 12(5), 535-540

Kumari, V., Checkley, S. A., & Gray, J. A. (1996). Effect of cigarette smoking on prepulse inhibition of the acoustic startle reflex in healthy male smokers. Psychopharmacology, 128(1), 54-60.

Lamm, C., Decety, J., Singer, T., 2011. Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. Neuroimage 54, 2492–502.

MacDonald, A.W., 2000. Dissociating the Role of the Dorsolateral Prefrontal and Anterior Cingulate Cortex in Cognitive Control. Science 288, 1835–1838.

Machlin SR, Harris GJ, Pearlson GD, Hoehn-Saric R, Jeffery P, Camargo EE (1991) Elevated medial-frontal cerebral blood flow in obsessive-compulsive patients: a SPECT study. Am J Psychiatry 148:1240-1242.

Machlin, S.R., Harris, G.J., Pearlson, G.D., Hoehn-Saric, R. (1991. Elevated medial-frontal cerebral blood flow in obsessive-compulsive patients: A SPECT study. The American Journal of Psychiatry 148, 1240-42.

Magno E, Foxe JJ, Molholm S, Robertson IH, Garavan H (2006) The anterior cingulate and error avoidance. J Neurosci 26:4769-4773.

Magno, E., Foxe, J.J., Molholm, S., Robertson, I.H., Garavan, H., 2006. The anterior cingulate and error avoidance. J Neurosci 26, 4769–4773.

Mansouri, F. A., Buckley, M. J., & Tanaka, K. (2007). Mnemonic function of the dorsolateral prefrontal cortex in conflict-induced behavioral adjustment. Science, 318(5852), 987-990.

McClure, S. M., Li, J., Tomlin, D., Cypert, K. S., Montague, L. M., & Montague, P. R. (2004). Neural correlates of behavioral preference for culturally familiar drinks. Neuron, 44(2), 379-387.

Miller, E.K., Cohen, J.D., 2001. An integrative theory of prefrontal cortex function. Annu. Rev. Neurosci. 24, 167–202.

Müller, N. G., & Knight, R. T. (2006). The functional neuroanatomy of working memory: contributions of human brain lesion studies. Neuroscience, 139(1), 51-58.

Nakamura, K., Roesch, M. R., & Olson, C. R. (2005). Neuronal activity in macaque SEF and ACC during performance of tasks involving conflict. Journal of neurophysiology, 93(2), 884-908.

Naqvi, N. H., Rudrauf, D., Damasio, H., & Bechara, A. (2007). Damage to the insula disrupts addiction to cigarette smoking. Science, 315(5811), 531-534.

Nee DE, Kastner S, Brown JW (2011) Functional heterogeneity of conflict, error, task-switching, and unexpectedness effects within medial prefrontal cortex. Neuroimage 54:528-540.

Nee, D. E., & Jonides, J. (2008). Neural correlates of access to short-term memory. Proceedings of the National Academy of Sciences, 105(37), 14228-14233.

Nee, D.E., Kastner, S., Brown, J.W., 2011. Functional heterogeneity of conflict, error, task-switching, and unexpectedness effects within medial prefrontal cortex. Neuroimage 54, 528–40.

Nelson, J. K., Reuter-Lorenz, P. A., Sylvester, C. Y. C., Jonides, J., & Smith, E. E. (2003). Dissociable neural mechanisms underlying response-based and familiarity-based conflict in working memory. Proceedings of the National Academy of Sciences, 100(19), 11171-11175

Newman SD, Greco JA, Lee D (2009) An fMRI study of the Tower of London: a look at problem structure differences. Brain Res 1286:123-132.

Nieuwenhuis, S., Forstmann, B.U., Wagenmakers, E.-J., 2011. Erroneous analyses of interactions in neuroscience: a problem of significance. Nat. Neurosci. 14, 1105–1107.

O'Reilly, J.X., Schüffelgen, U., Cuell, S.F., Behrens, T.E.J., Mars, R.B., Rushworth, M.F.S., 2013. Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. Proc. Natl. Acad. Sci. U. S. A. 110.

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. Science, 304(5669), 452-454.

Oliveira, F.T., McDonald, J.J., Goodman, D., 2007. Performance monitoring in the anterior cingulate is not all error related: expectancy deviation and the representation of action-outcome associations. J. Cogn. Neurosci. 19, 1994–2004.

Paulus, M. P., & Frank, L. R. (2006). Anterior cingulate activity modulates nonlinear

Paus, T., 2001. Primate anterior cingulate cortex: where motor control, drive and cognition interface. Nat. Rev. Neurosci. 2, 417–424.

Petry NM, Casarella T (1999) Excessive discounting of delayed rewards in substance abusers with gambling problems. Drug Alcohol Depend 56:25-32.

Plassmann, H., Kenning, P., Deppe, M., Kugel, H., & Schwindt, W. (2008). How choice ambiguity modulates activity in brain areas representing brand preference: evidence from consumer neuroscience. Journal of Consumer Behaviour, 7(4-5), 360-367.

Preuschoff, K., Quartz, S.R., Bossaerts, P., 2008. Human insula activation reflects risk prediction errors as well as risk. J. Neurosci. 28, 2745–52.

Rainville, P., Duncan, G. H., Price, D. D., Carrier, B., & Bushnell, M. C. (1997). Pain affect encoded in human anterior cingulate but not somatosensory cortex. Science, 277(5328), 968-971.

Rolls, E. T. (2004). The functions of the orbitofrontal cortex. Brain and cognition, 55(1), 11-29.

Rushworth MF, Buckley MJ, Behrens TE, Walton ME, Bannerman DM (2007) Functional organization of the medial frontal cortex. Current opinion in neurobiology 17:220-227.

Rushworth, M. F. S., & Behrens, T. E. J. (2008). Choice, uncertainty and value in prefrontal and

Rushworth, M. F. S., Walton, M. E., Kennerley, S. W., & Bannerman, D. M. (2004). Action sets

Saal, D., Dong, Y., Bonci, A., & Malenka, R. C. (2003). Drugs of abuse and stress trigger a common synaptic adaptation in dopamine neurons. Neuron, 37(4), 577-582.

Scheffers, M. K., Coles, M. G., Bernstein, P., Gehring, W. J., & Donchin, E. (1996). Event-related brain potentials and error-related processing: An analysis of incorrect responses to go and no-go stimuli. Psychophysiology, 33(1), 42-53.

Scheffers, M.K., Coles, M.G.H., 2000. Performance Monitoring in a Confusing World : Error-Related Brain Activity, Judgments of Response Accuracy, and Types of Errors. J. Exp. Psychol. 26, 141–151.

Scheffers, M.K., Coles, M.G.H., 2000. Performance Monitoring in a Confusing World : Error-Related Brain Activity, Judgments of Response Accuracy, and Types of Errors. J. Exp. Psychol. 26, 141–151.

Schoenbaum G, Setlow B, Saddoris MP, Gallagher M (2003) Encoding predicted outcome and acquired value in orbitofrontal cortex during cue sampling depends upon input from basolateral amygdala. Neuron 39:855-867.

Schönberg, T., Daw, N. D., Daphna, J., & O'Doherty, J. P. (2007). Reinforcement learning signals in the human striatum distinguish learners from nonlearners during reward-based decision making. The Journal of Neuroscience, 27(47), 12860-12867.

Schultz, W., & Dickinson, A. (2000). Neuronal coding of prediction errors. Annual Review of Neuroscience, 23, 473-500.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. Science, 275(5306), 1593-1599.

Seamans, J. K., & Yang, C. R. (2004). The principal features and mechanisms of dopamine modulation in the prefrontal cortex. Progress in Neurobiology, 74, 1–57.

Serences, J.T., 2004. A comparison of methods for characterizing the event-related BOLD timeseries in rapid fMRI. Neuroimage 21, 1690–700.

Shackman, A.J., Salomons, T. V, Slagter, H. a, Fox, A.S., Winter, J.J., Davidson, R.J., 2011. The integration of negative affect, pain and cognitive control in the cingulate cortex. Nat. Rev. Neurosci. 12, 154–67.

Shadmehr R, Wise SP (2004) Motor Learning and Memory for Reaching and Pointing. In: The Cognitive Neurosciences III, 3rd Edition (Gazzaniga M, ed). Cambridge: MIT Press.

Shidara, M., Richmond, B.J., 2002. Anterior cingulate: Single neuronal signals related to degree of reward expectancy. Science (80-.). 296, 1709–1711.

Shima, K., Tanji, J., 1998. Role for cingulate motor area cells in voluntary movement selection based on reward. Science 282, 1335–1338.

Siegel, S. (1999). Drug anticipation and drug addiction. The 1998 H. David Archibald lecture. Addiction, 94(8), 1113-1124.

Silvetti, M., Seurinck, R., Verguts, T., 2011. Value and prediction error in medial frontal cortex: integrating the single-unit and systems levels of analysis. Front. Hum. Neurosci. 5, 75.

Sohn MH, Albert MV, Jung K, Carter CS, Anderson JR (2007) Anticipation of conflict monitoring in the anterior cingulate cortex and the prefrontal cortex. PNAS, 104:10330-10334.

Sporns, O., Tononi, G., & Kötter, R. (2005). The human connectome: a structural description of the human brain. PLoS computational biology, 1(4), e42.

Steinhauser, M., Maier, M., Hübner, R., 2008. Modeling behavioral measures of error detection in choice tasks: response monitoring versus conflict monitoring. J. Exp. Psychol. 34, 158–76.

Stephens, D. W. & Krebs, J. R. (1986). Foraging Theory. Princeton, NJ: Princeton University Press.

Stewart, J., & Eikelboom, R. (1987). Conditioned drug effects. In Handbook of psychopharmacology (pp. 1-57). Springer US.

Stigler, G. J. (1965). Essays in the History of Economics (Vol. 3). Chicago: University of Chicago Press.

Suhr, J. A., & Tsanadis, J. (2007). Affect and personality correlates of the Iowa Gambling Task. Personality and Individual Differences, 43(1), 27-36.

Sutton RS, Barto AG (1998) Reinforcement Learning. Cambridge: MIT Press.

Swick, D., & Turken, U. (2002). Dissociation between conflict detection and error monitoring in the human anterior cingulate cortex. Proceedings of the National Academy of Sciences, 99(25), 16354-16359.

Taren, A. a, Venkatraman, V., Huettel, S. a, 2011. A parallel functional topography between medial and lateral prefrontal cortex: evidence and implications for cognitive control. J. Neurosci. 31, 5026–31.

Thorndike EL (1911) Animal intelligence: Experimental studies. New York: MacMillan.

Tremblay L, Schultz W (1999) Relative reward preference in primate orbitofrontal cortex. Nature 398:704-708.

Tremblay, L., & Schultz, W. (1999). Relative reward preference in primate orbitofrontal cortex. nature, 398(6729), 704-708.

Valentin VV, Dickinson A, O'Doherty JP (2007) Determining the neural substrates of goal-directed learning in the human brain. J Neurosci 27:4019-4026.

van der Meer MA, Redish AD (2010) Expectancies in decision making, reinforcement learning, and ventral striatum. Front Neurosci 4:6.

van Veen, V. V., & Carter, C. S. (2002b). The timing of action-monitoring processes in the anterior cingulate cortex. Journal of Cognitive Neuroscience, 14(4), 593-602.

van Veen, V., & Carter, C. S. (2002a). The anterior cingulate as a conflict monitor: fMRI and ERP studies. Physiology and Behavior, 77(4), 477-482.

van Veen, V., Cohen, J. D., Botvinick, M. M., Stenger, V. A., & Carter, C. S. (2001). Anterior cingulate cortex, conflict monitoring, and levels of processing. Neuroimage, 14(6), 1302-1308

Verdejo-García, A., Bechara, A., Recknor, E. C., & Pérez-García, M. (2007). Negative emotion-driven impulsivity predicts substance dependence problems. Drug and alcohol dependence, 91(2), 213-219.

Walton, M.E., Devlin, J.T., Rushworth, M.F.S., 2004. Interactions between decision making and performance monitoring within prefrontal cortex. Nat Neurosci 7, 1259–1265.

Williams, S. M., & Goldman-Rakic, P. S. (1998). Widespread origin of the primate mesofrontal dopamine system. Cerebral Cortex, 8(4), 321-345.

Xue, G., Lu, Z., Levin, I. P., & Bechara, A. (2010). The impact of prior risk experiences on subsequent risky decision-making: the role of the insula. Neuroimage, 50(2), 709-716.

Xue, G., Lu, Z., Levin, I. P., Weller, J. A., Li, X., & Bechara, A. (2009). Functional dissociations

Yarkoni, T. (2009). Big correlations in little studies: Inflated fMRI correlations reflect low statistical power—Commentary on Vul et al.(2009). Perspectives on Psychological Science, 4(3), 294-298.

Yarkoni, T., Poldrack, R. a, Nichols, T.E., Van Essen, D.C., Wager, T.D., 2011. Large-scale automated synthesis of human functional neuroimaging data. Nat. Methods 8, 665–70.

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: conflict monitoring and the error-related negativity. Psychological review, 111(4), 931.

Yeung, N., Botvinick, M.M., Cohen, J.D., 2004. The Neural Basis of Error Detection : Conflict Monitoring and the Error-Related Negativity. Psychol. Rev. 111, 931–959.

Yeung, N., Nieuwenhuis, S., 2009. Dissociating response conflict and error likelihood in anterior cingulate cortex. J. Neurosci. 29, 14506–10.

Andrew Jahn, Ph.D.
Curriculum Vitae
May 2015

Address:
Indiana University
Department of Psychological & Brain Sciences
1101 E. 10th St
Room A216C
Bloomington, IN 47405

Work Phone: 812.856.1846
Home Phone: 952.769.7126
email: ajahn@indiana.edu

EDUCATION

2015-present	Postdoctoral fellow, Haskins Laboratories, New Haven, CT
2010-2015	Ph.D., Psychology/Cognitive Neuroscience, Indiana University, IN
2004-2008	B.A., Psychology (Magna Cum Laude), Carleton College, Northfield, MN

PUBLICATIONS

Refereed Journal Articles

1. Mendizabal, N. V., Jones, D. R., **Jahn, A.,** Bies, R. R., & Brown, J. W. (2014). Nicotine and Cotinine Exposure from Electronic Cigarettes: A Population Approach. *Clinical Pharmacokinetics.* Advance online publication. doi: 10.1007/s40262-014-0221-7

2. **Jahn, A.,** Nee, D. E., Alexander, W. H., & Brown, J. W. (2014). Distinct regions of anterior cingulate cortex signal prediction and outcome evaluation. *Neuroimage,* 95, 80-89.

3. Nee, D. E., **Jahn, A.,** & Brown, J. W. (2014). Prefrontal Cortex Organization: Dissociating Effects of Temporal Abstraction, Relational Abstraction, and Integration with fMRI. *Cerebral Cortex,* 24, 2377-87.

4. **Jahn, A.,** Nee, D. E., & Brown, J. W. (2011). The neural basis of predicting the outcomes of imagined actions. *Frontiers in Decision Neuroscience,* 5:128.

5. Mowrer, S. M., **Jahn, A.,** Cunningham, W. A., & Abduljalil, A. M. (2011). The value of success: Acquiring gains, avoiding losses, and simply being successful. *PLoS One,* 6:9.

6. Cunningham, W. A., Arbuckle, N. L, **Jahn, A.,** Mowrer, S. M., & Abduljalil, A. M. (2010). Aspects of neuroticism and the amygdala: Chronic tuning from motivational styles. *Neuropsychologia,* 49, 657-662.

Book Chapters

1. Cunningham, W. A., Johnsen, I. R., & **Jahn, A.** (2011). Attitudes. In J. Decety & J. Cacioppo (Eds.), The Handbook of Social Neuroscience. Oxford, UK: Oxford University Press.

Manuscripts in Preparation

1. **Jahn, A.,** Nee, D. E., Alexander, W. H., & Brown, J. W. (in preparation for *Neuroimage*). Pain, Incongruency, and Surprise: Prediction violation across domains in the anterior cingulate cortex.

2. Cunningham, W. A., Mowrer, S. M., **Jahn, A.,** & Kesek, A. K. (in prep). Dissociating decision related from anticipatory activity in orbitofrontal cortex.

3. Cunningham, W. A., Van Bavel, J. J., & **Jahn, A.** (in prep). Attention and the neural components of affective processing.

CONFERENCE PRESENTATIONS

Posters

2014 **Jahn, A.,** Strait, C., Brown, J. W., & Hayden, B. Testing computational models of anterior cingulate cortex with monkey single units. Poster session to be presented at: Annual Meeting of Society for Neuroscience; November 2014; Washington, D.C.

2014 **Jahn, A.,** Nee, D. E., Alexander, W. H., Brown, J. W. Medial prefrontal cortex signals prediction errors across multiple domains of pain and cognitive control. Poster session to be presented at: Annual Meeting of Society for Neuroscience; November 2014; Washington, D.C.

2013 **Jahn, A.,** Nee, D. E., Alexander, W., & Brown, J. W. Distinct regions of anterior cingulate cortex signal prediction and outcome evaluation. Poster session presented at: Annual Meeting of Society for Neuroscience; November 2013; San Diego, CA.

2013 **Jahn, A.,** Nee, D. E., Alexander, W., & Brown, J. W. Pain, Congruency, and Surprise: Prediction Violation across Domains in the Anterior Cingulate Cortex. Poster session presented at: Annual Meeting of Cognitive Neuroscience Society; April 2013; San Francisco, CA.

2012 **Jahn, A.,** Nee, D. E., Alexander, W., & Brown, J. W. Distinct regions of anterior cingulate cortex signal prediction and outcome evaluation. Poster session presented at: Annual Midwestern Cognitive Science Conference; May 2012; Bloomington, IN.

2012	**Jahn, A.,** Nee, D. E., Alexander, W., & Brown, J. W. Distinct regions of anterior cingulate cortex signal prediction and outcome evaluation. Poster session presented at: Annual Meeting of Cognitive Neuroscience Society; April 2012; Chicago, IL.
2011	**Jahn, A.,** Nee, D. E., & Brown, J. W. The neural basis of predicting the outcomes of planned actions. Poster session presented at: Annual Meeting of Society for Neuroscience; November,2011; Washington, D.C.
2010	Mowrer, S. M., **Jahn, A.,** Cunningham, W. A., & Abduljalil, A. M. Is obtaining a reward different from avoiding an aversive outcome? Separable effects of stimulus and outcome evaluation and their interaction. Poster session presented at: Annual Meeting of Social and Affective Neuroscience; October 2010; Chicago, IL.
2010	Arbuckle, N., Mowrer, S.M., **Jahn, A.,** Abduljalil, A. M., & Cunningham, W. A. Aspects of neuroticism and the amygdala: Chronic tuning from motivational style. Poster session presented at: Annual Meeting of Social and Affective Neuroscience; October 2010; Chicago, IL.

TEACHING EXPERIENCE

Teaching Interests: Cognitive neuroscience, neuroimaging methods, univariate and multivariate statistics, Bayesian statistics, computational modeling, experimental design, decision-making

Primary Instructor

2013-present	Introductory Statistics
2013-present	Introductory Psychology

Lab Instructor

2012	Methods of Experimental Psychology
2011	Introduction to Neuroimaging Methods

Private Tutor

2014-present	fMRI Data Analysis Tutor via Google Helpouts
2012-present	Univariate and Multivariate Statistics
2012-2013	fMRI Methods

Freelance

2014	Neuroimaging Advisor (fMRI Experiment Design): Ohio State University
2013-2014	Neuroimaging Advisor (PET data analysis): Michigan State University

Web-Based, Open-Access

2012-present andysbrainblog.blogspot.com (fMRI data analysis, computational modeling, and statistics tutorials, along with associated YouTube tutorials)

RESEARCH EXPERIENCE

2013 Visiting Scholar for Neurophysiological Methods: Rochester University

2010-2011 Imaging Research Facility Research Assistant: Indiana University

2008-2010 Research Assistant and Lab Manager: Ohio State University

2007-2008: Research Assistant: Carleton College

INVITED LECTURES

"Empirical comparison of computational models of mPFC function." University of Rochester, NY. 17 July, 2013.

PROFESSIONAL SERVICE

Ad Hoc Reviewer: *Neuroimage; Cognitive, Affective, and Behavioral Neuroscience; Frontiers in Human Neuroscience; Journal of Cognitive Neuroscience*

UNIVERSITY SERVICE

2013-2014 IU Cognitive Neuroscience Seminar Coordinator

2010-2011 IU Neuroimaging Group Coordinator

EXTRACURRICULAR UNIVERSITY SERVICE

2012-present Piano accompanist: Jacobs School of Music, Strings Department

RELATED PROFESSIONAL SKILLS

Neuroimaging Software AFNI, SPM, FSL
Statistical Software Matlab, R, SPSS
Stimulus Presentation E-Prime, Presentation
Programming Languages Python, AWK, C

PROFESSIONAL MEMBERSHIPS

Sigma Xi
Society for Neuroscience
Cognitive Neuroscience Society